DOCUMENT RESUME

ED 284 403 EC 200 221

AUTHOR	Hofmeister, Alan M.
TITLE	Assessing the Accuracy of a Knowledge-Based System: Special Education Regulations & Procedures. Final Report.
INSTITUTION	Utah State Univ., Logan.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE	86
GRANT	G008530236
NOTE	242p.; For related document, see EC 200 220. Some charts in appendices contain small print and will not reproduce well.
PUB TYPE	Reports - Research/Technical (143) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	*Artificial Intelligence; Computer Software; *Computer Uses in Education; Decision Making; *Disabilities; Elementary Secondary Education; *Expert Systems; *Individualized Education Programs; *Special Education; Summative Evaluation; Systems Development
IDENTIFIERS	Mandate Consultant (Computer System)

ABSTRACT

The purpose of this research project was the development and initial validation of Mandate Consultant, an expert system that provides a second opinion of the appropriateness of the decision-making process used in the development of Individualized Education Programs with handicapped children. (Expert systems are a development of artificial intelligence that combines the computer's capacity for storing specialized knowledge with a general set of rules intended to replicate the decision-making process of a human expert.) This expert system was designed as a tool for administrators of special education programs. The project utilized a four-stage product development approach: (1) product definition; (2) product design; (3) prototype development; and (4) validation of product performance. The validation consisted of a summative evaluation designed to assess whether the expert system accurately emulated the knowledge of experienced educators. It was found that conclusions produced by consultations using Mandate Consultant generally matched the conclusions of the "better" human experts and exceeded the conclusions of the majority of human experts. Thirty-five appendixes comprising more than half the document describe the work of all four phases of the development of Mandate Consultant. (KM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

MAY 7 1988

ED 1.310/2:284403

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED284403

FINAL REPORT

Assessing the Accuracy of a Knowledge - Based System: Special Education Regulations & Procedures

Special Education Program
Department of Education
Grant # G008530236
Project # 023BH50056

Project Director:
Alan M. Hofmeister
Utah State University
1986

ORIGINAL

BEST COPY AVAILABLE

TABLE OF CONTENTS

Page

LIST OF TABLES iv

LIST OF FIGURES vi

INTRODUCTION 1

Implications of the Law 1
Concerns Associated With Hearings 2
The Need 5
A Computer-Based Expert System as a Viable Solution 6
Problem 9
Purpose 10
Objectives 10
Research Questions 10

PROCEDURES 12

Product Definition 12
Product Design 18
Product Prototype 27
Product Validation 36

RESULTS AND DISCUSSION 58

Results 58
Discussion 79

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS 89

Summary 89
Conclusions 93
Recommendations 95

REFERENCES 97

APPENDICES 107

Appendix A. Summary of Studies Addressing the Evaluation of Computer-based Expert Systems 108

TABLE OF CONTENTS (continued)

page

Appendix B. An Analysis of the Studies Addressing Computer-based Expert Systems 111
Appendix C. Product Definition Appraisal Plan 113
Appendix D. Description of Mandate Consultant 115
Appendix E. Evaluation Preview for Mandate Consultant 119
Appendix F. Questionnaire for Product Definition Phase 125
Appendix G. Validity and Reliability of Product Definition Questionnaire 128
Appendix H. Product Design Appraisal Plan 131
Appendix I. Clarity Evaluation Instrument for Product Design Phase 133
Appendix J. Validity and Reliability of Clarity Evaluation Instrument 137
Appendix K. Sample of Mandate Consultant Program Code 139
Appendix L. Example of Expert System Output 144
Appendix M. Internal Consistency Worksheet 151
Appendix N. Regulations Checklist for Product Design Phase 154
Appendix O. Interview Form for Product Design Phase 160
Appendix P. Validity and Reliability of Internal Consistency Worksheet for Product Design Phase 162
Appendix Q. Validity and Reliability of Regulation Checklist for the Product Delevopment Phase 164
Appendix R. Validity and Reliability of Interview Form for Product Design Phase 167
Appendix S. Prototype Appraisal Plan 169
Appendix T. Documentation for Operating Mandate Consultant: An Expert System for Application in Special Education 171
Appendix U. Test Case Data Collection Form 185
Appendix V. Product Prototype Questionnaire 188
Appendix W. Validity and Reliability of the Questionnaire for the Product Prototype Phase 193
Appendix X. Sample Letter to Cache County School District Special Educators 195
Appendix Y. Letter/Directions for Human Experts of Phase I 198
Appendix Z. Cumulative File Report Form 202

TABLE OF CONTENTS (continued)

page

Appendix AA. Informed Consent Form for Human Experts 205
Appendix BB. General Information Form for Human Experts 208
Appendix CC. Suggestion Notice 210
Appendix DD. Letter/Directions for Evaluators of Phase II 211
Appendix EE. Informed Consent Form for Evaluators 214
Appendix FF. General Information Form for Evaluators 216
Appendix GG. Rating Form for Phase II 218
Appendix HH. Expert Conclusions Coded for Each Case by Special Education Graduate Student 220
Appendix II. Independent Evaluator Ratings for Each Expert's Conclusions 226

LIST OF TABLES

Table Page

2. Phase I: Human Experts' Qualifications 41

3. Phase II: Evaluator's Qualifications 42

4. Number of Mandate Consultant (MC) and Human Expert (HE) Conclusions 60

5. Number of Interexpert Agreements (A) and Disagreements (D) for Mandate Consultant (MC) and Six Human Experts (HE) 51

6. Correlation Matrix: Pearson r's for Each Pair of Experts Based on Number of Interexpert Agreements 62

7. Number of Most Frequently-Noted Conclusions By Mandate Consultant (MC) and Human Experts (HE) . . 63

8. Rankings of Experts by Total Conclusions, Interexpert Agreements, and Most Frequently-Noted Conclusions 66

9. Results of Kendall's Coefficient of Rank Correlation (Tau) 67

10. Interevaluator Reliability Coefficients for Interevaluator Ratings 69

11. Evaluator Ratings of Mandate Consultant (MC) and Human Expert (HE) Case Reports by Evaluator . . . 71

12. Rank Ordering of Mandate Consultant (MC) and Human Expert (HE) Based on Percent of Reports Judged as Acceptable 72

13. Kendall's Coefficient of Concordance Correlation (W) Results: Three Sets of Evaluator Rankings for Seven Experts 73

14. Percentage of Expert Case Reports Judged Acceptable or Unacceptable 75

15. Reliability Measures for Phase I Users of Mandate Consultant Reliability Coefficients . . . 79

LIST OF TABLES (continued)

Table page

16. Summary of Studies Addressing the Evaluation of Computer-based Expert Systems 109

LIST OF FIGURES

Figure Page

1. Research and Development Model for Mandate Consultant 13

2. Design for Product Validation 36

3. Totals by Expert Across the Ten Cases 65

4. Percentage of Expert Case Reports Judged Acceptable 76

5. An Analysis of the Studies Addressing Computer-based Expert Systems 112

6. Product Definition Appraisal Plan 114

7. Flowchart: How Mandate Consultant Operates . . . 117

8. Product Design Appraisal Plan 132

9. Prototype Appraisal Plan 170

10. Flowchart: How Mandate Consultant Operates . . . 175

INTRODUCTION

Implications of the Law

During the past several decades, general acceptance of public education of handicapped children in the United States has grown rapidly (Henderson & Hage, 1979). This general acceptance led to the enactment of the Education for All Handicapped Children Act of 1975 (Smith, 1981). Three of the key technical terms appearing in the Act include: "free appropriate public education" (FAPE), "least restrictive environment" (LRE), and "individualized education program" (IEP) (Zirkel, 1983). The Act assures that all children have available to them a free, appropriate, public education. To the maximum extent appropriate, handicapped children are to be educated with children who are not handicapped, that is, educated in the least restrictive environment. The written IEP identifies the plan for meeting the FAPE and LRE requirements. A team that includes school representatives and parents meets to develop the IEP.

The IEP "process" serves as a forum within which parents and school officials meet to reach agreement on the content and provision of the handicapped child's education. When school officials and parents cannot agree on an educational plan for the child, various types of mediation are usually attempted. Low-level mediation might involve ongoing discussions between various local school officials and parents that, hopefully, produce compromise and agreement of an appropriate educational plan. High-level mediation might involve a third party from the state education agency to negotiate a settlement between school personnel and parents (Lehr & Haubrich, 1986). If a settlement is not reached, an impartial due process hearing officer acts as an arbitrator to achieve resolution (Kammerlohr, Henderson, & Rock, 1983).

The due process provisions of the Act were designed to protect the rights of parents and the handicapped child. Under the Act, parents become partners in the educational decision-making process. The interaction between school officials and parents usually results in agreement on educational programs. In other instances, relationships are strained, and mediation is unsuccessful (Fiedler, 1985). Parents or school officials can request a hearing officer to resolve the situation. The hearing officer reviews documents (evaluations, progress reports, etc.) from the child's file and listens to testimony from the involved parties. The officer makes a ruling based on the Act and state regulations.

Concerns Associated With Hearings

Delays Accompany Hearings

There are problems associated with the implementation of due process hearings. First, hearings cause delays and interruptions in appropriate services for handicapped students (Budoff, Orenstein, & Abramson, 1981). The Act states that services for the handicapped child should be maintained during the hearing process. Nonetheless, hearing officers have other obligations. Thus, scheduling due process hearings to limit, if not avoid, delay can be difficult (Smith, 1981). While the handicapped child may continue to receive services during delays, the disputed services may not be appropriate.

Hearings Are Costly

A second major concern is cost. A cost analysis conducted in the late 1970s showed that the typical due process hearing cost between $2,000 and $4,000 (Henderson & Hage, 1979). The school district bears this cost regardless of whether parents or school officials request the hearing. The $4,000 figure approaches the current average cost for educating a handicapped student annually; also, this figure is nearly twice the current average cost for educating a nonhandicapped student (Belsches-Simmons & Lines, 1984). From this perspective, a more cost-effective means of dealing with differences between parents of handicapped children and school officials would be desirable (Kammerlohr et al., 1983).

In addition to the monetary costs associated with due process hearings, costs accrue in terms of the stress on the parties involved. Involvement in the due process hearing often establishes an attitudinal and communicative barrier between parents and school officials. Unfortunately, the stress associated with the hearing frequently carries over into the relationships between the parent and school officials following the hearing (Fiedler, 1985). Keeping the student immune to the effects of the stress presents a difficult, if not impossible, task for both the parents and school personnel.

Hearings Do Not Ensure Equitable Decisions

A third major problem associated with due process hearings relates to an underlying requirement of due process, that is, the hearing officer must be impartial and competent (Salend & Zirkel, 1984). Although regulations specifically state that a hearing officer may not be an employee of the local school district involved in the case (Ekstrand, 1979), the regulations provide no other specific criteria for determining impartiality.

In addition, federal regulations do not specify the background or training necessary for hearing officers (Smith, 1981). Thus, the background experiences of hearing officers vary greatly (Ekstrand, 1979). In a study of hearing officers in North Carolina, the diversified professions included a homemaker, postmaster, and research biologist (Turnbull, Strickland, & Turnbull, 1981). However, the diversity of professions and past experiences do not weaken the process as much as the lack of training, or more appropriately, the disparity of training for hearing officers (Turnbull, Turnbull, & Strickland, 1979). Generally, the staff of the state education agency are responsible for training their state's hearing officers. While ongoing efforts to meet this responsibility have occurred, these efforts vary greatly from state to state (Smith, 1981). Unfortunately, state agency staff have not consistently attended to competence and impartiality in the selection of hearing officers. Thus current due process hearings do not ensure equitable and effective educational decisions (Salend & Zirkel, 1984).

The Need

The author's review of the literature clearly documents problems associated with many due process hearings: (a) Disagreement can delay services to students; (b) hearings are costly in terms of finances and stress; and (c) the knowledge and skill of hearing officers vary greatly. Clearly, these problems substantiate the need to resolve disagreements between school officials and parents prior to formal due process hearings (Belsches-Simmons & Lines, 1984). School officials and parent advocates need an unbiased, knowledgeable expert to objectively review problems regarding development of an educational program for a handicapped child. The consultant should provide objective feedback that can be used as a basis for decision making. In addition, from the perspective of the school officials, such a consultant should be readily available at a reasonable cost (Parry & Ferrara, 1985).

A Computer-Based Expert System as a Viable Solution

In recent years the field of artificial intelligence (AI) and applications of AI, such as expert systems, have gained much media attention. But what is artificial intelligence?

> Artificial intelligence is the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behavior--understanding, language, learning, reasoning, solving problems, and so on. (Barr & Feigenbaum, 1981, p. 3)

Artificial intelligence systems intended to replicate decision making by knowledgeable and experienced humans are called expert systems. The developer typically designs an expert system to engage the user in a dialogue. This dialogue in many ways parallels the type of conversation a person might have with an expert consultant. The computer is programmed to ask the user questions to clarify the problem or situation (Barr & Feigenbaum, 1981). For example, MYCIN is a well-known medical system for physicians (Davis, Buchanan, & Shortliffe, 1975). With MYCIN, the user feeds into the computer information on the characteristics of the patient's symptoms. The computer is programmed to match the patient's data with information in the program on the characteristics of bacterial cultures and then, based on programmed logic, present a disease diagnosis.

Until recently there has been little application of expert system technology to the field of education (Hofmeister & Ferrara, in press). However, with the increased power and availability of computer hardware, and the gains in artificial intelligence technology, the development of expert systems for educators is now feasible.

Developing the knowledge base for expert systems often helps clarify complex bodies of knowledge. This clarification is essential to accurate logic representation within the knowledge base, but it also carries an associated benefit of disclosing gaps in our understanding of a subject domain (Duda & Shortliffe, 1983). Professionals, including educators, have previously avoided certain subject domains because the domains were viewed as unsuitable for formal organization. Representing the subject domain in an expert system knowledge base forces it to become systematized and codified. In addition, because of the attributes of computers and expert systems, the organized body of knowledge can produce consistently accurate information in a timely manner for application by professionals (Wills, Teather, & Innocent, 1982).

A considerable amount of educators' expertise has been gained through experience. Knowledge gained through experience can be described as heuristic knowledge (Harmon & King, 1985). Heuristic knowledge includes substantial, subjective information. Duda and Shortliffe (1983) indicate that clarifiying this subjective information into a knowledge base serves to increase its objectivity by making it more explicit and public. They state, "If the knowledge is valuable and is faithfully represented, the resulting program can make it more widely available and permit it to be more uniformly applied as an aide to decision-making" (p. 265).

In summary, few expert systems have been developed in the field of education. Based on the literature, only one expert system has been developed in the area of federal regulations and state rules for special education. Specifically, Ferrara and Hofmeister (1984) developed an expert system, "CLASS.LD," to provide a second opinion regarding the accuracy of placement decisions for learning disabled students. Using the system, educators can check their reasoning and placement conclusions with the decision rules of the expert system, based on federal and state regulations as well as research findings. Preliminary evaluation by the developers indicates that systems like CLASS.LD can (a) perform as well as humans in specific areas, and (b) clarify existing knowledge and identify areas where knowledge is needed (Hofmeister & Lubke, in press). Clearly, the potential remains great for expert systems that can assist in the interpretation and implementation of regulations (Waterman, 1986).

Faced with a relatively narrow class of information, that is, appropriate implementation of federal and state regulations regarding the development of IEPs, and recognizing the current problems associated with the due process hearings, development of an expert system emerges as a potential solution. As noted by Rychener (1985):

> The advantages of an expert system are significant enough to justify a major effort to build them. Decisions can be obtained more reliably and consistently. . . . A problem area can be standardized and formalized through the process of building an expert system for it. An expert system may be especially useful in a consultation mode on difficult cases, where humans may overlook obscure factors. An expert system can often serve as an example of a good strategy in approaching a problem, which might be useful in training situations. (pp. 30-31)

Problem

An understanding of recent technological advances suggests that development of an expert system to review the special education regulations, particularly those governing the develoment of an IEP, holds promise as a readily-available, unbiased, knowledgeable consultant for school officials and parents. To date, few such expert systems in the field of special education have been developed.

Purpose

The purpose of this project was the development and initial validation of a computer-based expert system that reviews appropriateness of the procedures followed for the development of IEPs. The expert system was designed as a tool for administrators of special education programs. The validation component of this study focused on establishing the accuracy of the knowledge base of the expert system.

Objectives

Specific objectives of this project were:

1. To determine whether it was feasible to emulate the knowledge of expert special education administrators through development of a computer-based expert system.
2. To determine if the knowledge base of a computer-based expert system was accurate as compared to the knowledge of expert special education administrators.

Research Questions

The research questions addressed by this dissertation were:

1. To what degree do expert system-generated conclusions match human expert conclusions in terms of noting discrepancies with special

education regulatory procedures for the development of IEPs?

2. To what degree are expert system-generated conclusions or human expert conclusions rated as acceptable (based on a rating scale of 1-ideal, 2-acceptable, 3-less than acceptable, and 4-unacceptable) by a panel of human expert evaluators?

PROCEDURES

The author of the study conducted a formative and a summative evaluation of Mandate Consultant (MC), an expert system developed primarily for educators. Specifically, Mandate Consultant was designed to assist educators with a review of the procedures used in development of a student's IEP. Borg and Gall (1983) advocated the use of an R&D model for the successful development and validation of educational products. An outline of the model selected for this product appears in Figure 1. Briefly, this product development approach consists of four stages: (a) definition of the product, (b) design of the product, (c) development of the prototype, and (d) validation of product performance.

Product Definition

The process of defining the product included the following components: (a) an appraisal plan identifying the major activities for this phase, (b) a review of the relevant literature, (c) an analysis of the problem and proposed solution, (d) identification of objectives for the development of the product, and (e) an appraisal of the activities completed during the definition phase. As a part of this phase, several documents were developed: a condensed version of the literature review, a brief description of the proposed expert system, and a preview of the evaluation

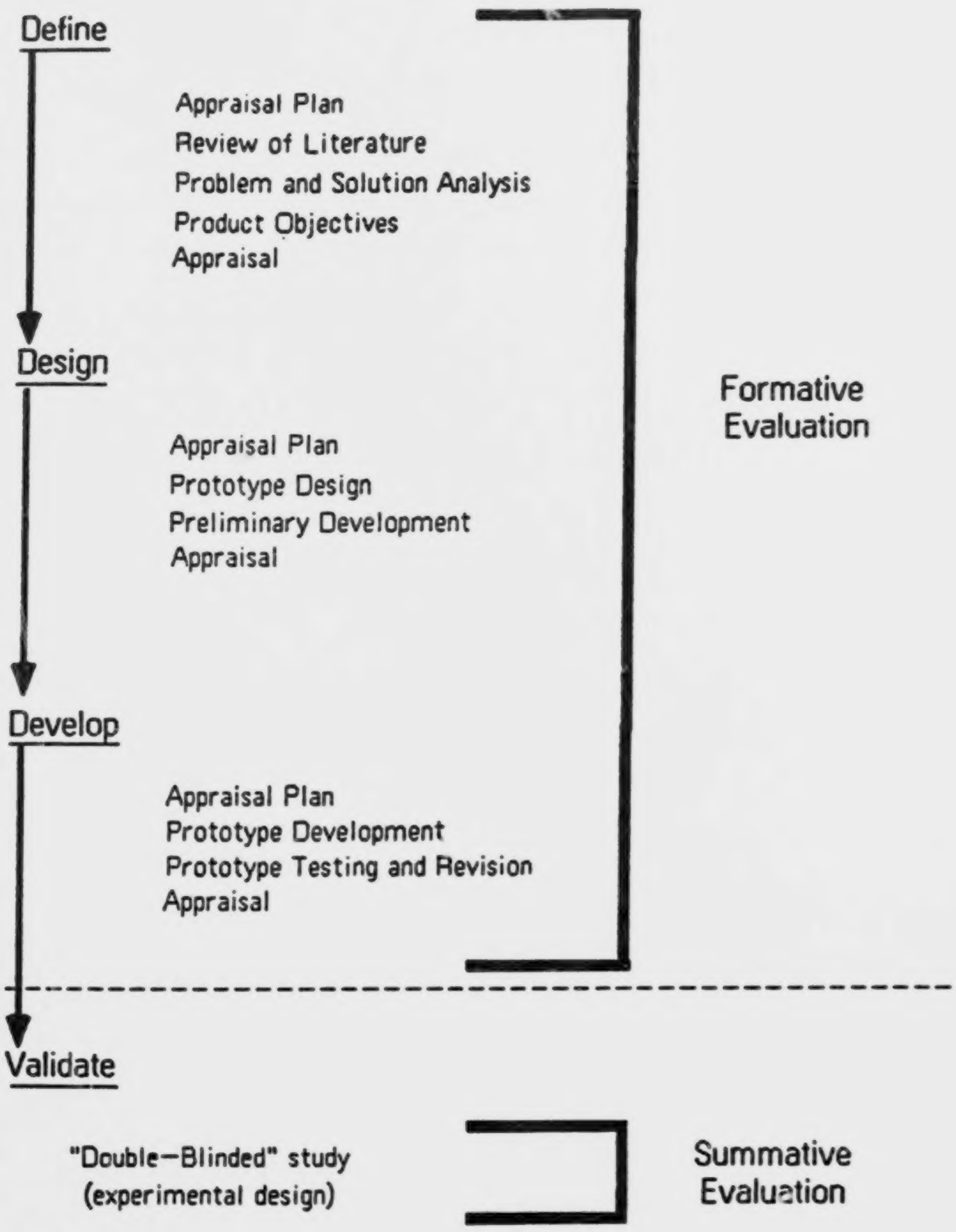


Figure 1. Research and Development Model for Mandate Consultant

planned for the development of the product. In addition, a product definition questionnaire was developed and implemented to gather the input of other special educators into the definition of the product. The various components of the product definition phase are presented in the following discussion.

Appraisal Plan

The first step to the definition phase involved developing an appraisal plan (see Appendix C) to provide structure for the developer. The plan operationalized the activities of the product definition phase. This systematic approach to development and evaluation activities was based on a model presented by Brinkerhoff, Brethower, Hluchyj, and Nowakowski (1983). The approach focused the developer's attention on relevant questions and procedures for completing and evaluating the definition phase.

Review of the Literature

The major questions for this phase dealt with establishing the need and the feasibility of the proposed solution. As a result, the author conducted a review of the literature to: (a) identify and evaluate existing research and related information addressing the implementation of the federal and state regulations in special education, (b) identify and evaluate existing research and related information regarding the application of expert system technology to the implementation of special education regulations, and (c) identify and describe the problem based on the literature review. The outcome of the review appears in the "Introduction" and "Review of Literature" chapters of this manuscript.

Problem and Solution Analysis

The developer of any product should initially clarify the problem addressed by the proposed development (Hood, 1973). In addition to the literature review for the proposed product, a group of professionals from special education were consulted to substantiate and clarify the problem and solution addressed by the proposed product. The group included: (a) a staff member from the Utah state education agency, (b) a due process hearing officer (trained in response to the mandate of PL 94-142), (c) a special education administrator from a local education agency, (d) a Utah State University (USU) faculty member from the special education department, and (e) a staff member of the USU Artificial Intelligence Research and Development Unit. Each read an eight-page condensed version of the literature review, a brief description of the proposed expert system (see Appendix D), and an evaluation preview stating the major steps and criteria for assessing product development (see Appendix E).

Product Definition Questionnaire

After reading the specified documents, each reviewer completed a data collection instrument entitled "Questionnaire for the Product Definition Phase" (see Appendix F). The questionnaire included twelve open-ended questions developed by the author. Responses to the questions provided clarification and substantiation of the problem and proposed solution.

Prior to implementing the questionnaire, the developer assessed the instrument's validity and reliability. According to Martuza (1977), a useful approach to the validation of a domain-referenced instrument involves ". . . (a) having two or more content specialists judge the relevance of each item to the objective it is intended to measure, and (b) using some index of interjudge agreement as the measure of item content validity" (p. 283). A technique based on this approach was implemented for estimating the validity of the questionnaire.

Specifically, two content specialists were asked to rate the relevance of each questionnnaire item to a respective objective using a four-point rating scale: (a) 1 = not relevant, (b) 2 = somewhat relevant, (c) 3 = quite relevant, and (d) 4 = very relevant (Martuza, 1977). Interrater agreement and an index of content validity were computed based on the specialists' responses. The outcome of the data analysis supported the instrument as a reliable measure and as validly representing its objectives. The raw data and results of the data analysis appear in Appendix G.

Results of the Product Definition Questionnaire

In general, the five professionals completing the questionnaire reached similar conclusions. First, the group concurred that there exists a clear need for the proposed expert system. All agreed that an expert system could provide assistance to special education administrators in reviewing their IEP development procedures. Some evaluators viewed the expert system as a relevant tool for parents, special educators in the classroom, and state agency staff as well. Second, the evaluators noted that, in general, the need for the proposed expert system appeared adequately documented and supported by data. However, some evaluators viewed the literature review as more comprehensive than others. Third, recognizing the constraints and resources of the proposed expert sytem, the evaluators viewed the expert system as a feasible solution for the identified problem.

Product Objective

The information generated by the literature review and the Product Definition Questionnaire provided the basis for the product development objective:

1. To develop an expert system designed to provide

school personnel with an expert review of the process used to develop the IEP of a selected student.

Appraisal

The appraisal plan (see Appendix C) included a criterion for evaluating the questionnaire activity of the product definition phase. Specifically, the responses of evaluators completing the Product Definition Questionnaire were compared with responses written by the developer. The agreement between evaluators' responses and the developer's responses exceeded the criterion of at least 80 percent agreement.

Product Design

The process of designing the product included the following components: (a) an appraisal plan identifying the major activities for this phase, (b) decisions regarding the initial prototype design, (c) preliminary development activities, and (d) appraisal of the activities of this phase. As part of the product design phase, two documents were developed: a listing of the proposed knowledge base rules of the expert system based on federal and state regulations, and examples of the output that would result from a consultation using the proposed expert system. In addition, several instruments were developed and implemented to gather input from other special educators for the design of the proposed product. These instruments included: (a) clarity evaluation instrument, (b) internal consistency worksheet, (c) regulations checklist, and (d) an interview form. The various components of this phase and the respective documents and instruments are discussed in the following sections.

Appraisal Plan

The purpose of the product design phase was to operationalize the product definition phase outcome into specific procedures for the expert system development process. During the design phase, a developer works to demonstrate the feasibility of developing the expert system (Gaschnig et al., 1983). To facilitate the design phase activities, the author formulated an appraisal plan similar to the one developed for the definition phase. This plan appears in Appendix H.

Prototype Design

A brief description and schema of the proposed expert system were developed (see Appendix D). The schema illustrated the possible infusion and operation of the proposed expert system into a school or parent advocacy environment. A group of special education professionals assessed the relationship between product objectives and the product design represented in the schema. Like the product definition phase, the group of professionals included: (a) a staff member from the Utah state education agency, (b) a due process hearing officer, (c) a special education administrator from a local education agency, (d) a Utah State University faculty member, and (e) a staff member of the Artificial Intelligence Research and Development Unit. Each read a condensed version of the literature review and the brief description of the proposed expert system (Appendix D).

Clarity Evaluation Instrument

After reading the specified documents, each reviewer completed a "Clarity Evaluation Instrument" (see Appendix I). The instrument included twelve objective-type items developed by the author. Responses to the items provided data to assess and clarify the proposed design.

Again, the author used the technique described by Martuza (1977) to estimate the validity and reliability of the instrument. Two content specialists reviewed the instrument and judged the relevance of each item to its respective objective, using the previously described four-point rating scale. Interrater agreement and an index of content validity were computed based on the specialists' responses. The outcome of the data analysis supported the instrument as a reliable measure, validly representing its objectives. The raw data and the results of the data analysis appear in Appendix J.

Results of the Clarity Evaluation Instrument

In general, the reviewers expressed similar judgments. First, they viewed the product objectives as clearly stated. Second, they agreed that the intent of the objectives was to design a product that provided an "expert" second opinion of IEP development procedures. Third, the reviewers judged the design schema (Appendix D) as a clear description of the proposed expert system operation. Finally, having reviewed the design schema, the evaluators concurred that the design includes (a) the anticipated primary users, (b) the data required to run the proposed system, and (c) the basis for the output generated by the system.

Preliminary Development

As part of the design phase, elements of the proposed expert system were conceptualized. First, 'if-then' statements, referred to as expert system rules, were written (see Appendix K). These rules, based on the federal and state regulations governing development of IEPs, comprised the "knowledge base" of the proposed expert system. Second, examples of the output to be generated by the proposed expert system were devised (see Appendix L). These examples illustrated the desired output from the proposed expert system using the knowledge base of rules.

Like earlier definition and design phases, professionals from the state education agency, local education agencies, and higher education assessed the preliminary development activities. As a part of these activities, the reviewers read a condensed literature review, a description of the proposed expert system, the rules from the knowledge base of the proposed system (see Appendix K), examples of the conclusions to be generated by the proposed system (see Appendix L), and the federal and state regulations governing the development of IEPs.

Preliminary Development Instruments

Instruments devised for collecting preliminary development data included: (a) an Internal Consistency Worksheet (see Appendix M), (b) a Regulations Checklist (see Appendix N), and (c) an Interview Form (see Appendix O). Completion of the Internal Consistency Worksheet involved two pencil and paper activities. The first activity once again assessed the relationship of the needs and corresponding objectives stated for the proposed system. The second activity required the reviewers to judge how likely the proposed product (based on the rules and the examples of output) might respond to the product development objectives.

Another essential element of the development process involved the review of the knowledge base rules by content experts (DeSalvo & Liebowitz, 1985). The purpose of this activity was to confirm that the knowledge base rules accurately represented the content of the federal and state regulations for special education. The author prepared a 41-item Regulations Checklist (see Appendix N) restating the content of the federal and state regulations governing the IEP development process. Three content specialists then compared the knowledge base rules with the items of the Regulation Checklist. Each specialist rated the items on the checklist using the following rating categories: (a) N = The knowledge base never addresses this regulation in the rules, (b) C = This regulation is cited but not used in the rules, (c) U = This regulation is used but not as fully intended, or (d) A = This regulation and the rules of the knowledge base agree totally.

The third instrument developed for the preliminary development phase was an Interview Form (see Appendix O). After local special education administrators and hearing officers reviewed examples of output to be generated from the proposed system, the author interviewed them. The Interview Form included three questions designed to evaluate the usefulness of the conclusions for the proposed primary users, school officials.

Once again, the Martuza (1977) technique was applied to estimate the validity and reliability of the three preliminary development instruments. Two content specialists reviewed each instrument and judged the relevance of each item to its respective objective using the previously described four-point rating scale. Interrater agreement and an index of content validity were computed based on the specialists' responses for each instrument. The author's interpretation of the results supported the instruments as reliable and valid measures of the respective objectives. The raw data and the results of the data analysis for the Internal Consistency Worksheet, the Regulations Checklist, and the Interview Form appear as Appendixes P, Q, and R, respectively.

Results of the Preliminary Development Instruments

Four reviewers completed the Internal Consistency Worksheet. The data provided strong support for the relationship between the needs identified in the literature review and preliminary product development based on the product objectives. The reviewers acknowledged the proposed system's use for generating a cost-effective, expert opinion regarding the regulations governing IEP development, but they also mentioned other potential benefits. For example, two reviewers listed the chronology with which the proposed system presents questions, and three listed the immediate feedback with specific citations from the regulations as additional benefits.

Three content specialists completed the Regulations Checklist after examining the knowledge base of the proposed system and reading the federal and state regulations. More than 91 percent of the reviewers' responses to checklist items indicated that the content of the regulations and the rules of the knowledge base agreed totally. Seven percent of the reviewers' responses indicated that the content of a regulation was addressed by the knowledge base rules but not as fully intended. Less than two percent of the reviewers responses indicated that the knowledge base rules failed to address specific regulations. The reviewers' responses served as the basis for several additions and changes to the knowledge base of the proposed system.

Finally, implementing the Interview Form developed for this phase, the author interviewed four educators. All reviewers perceived the output of the proposed system as potentially useful. However, they perceived that the system might not be used in practice. Their reasons varied greatly. Two of the reviewers believed administrators fail to acknowledge or accept the importance of following regulations when developing IEPs, and therefore, the system would not be used. The other two evaluators suggested that the output might not be used because the data the user fed into the proposed system could be invalid. Therefore, the "garbage in, garbage out" theory would apply and the output would be invalid. Two reviewers suggested that the output of the proposed system be extended to provide recommendations for solving or correcting the problems identified by the system. Another reviewer conjectured that solutions for the identified problems were implied in the immediate feedback, citing specific regulations. Finally, one reviewer suggested that subsequent development address additional qualitative or "best practice" concerns of IEP development. Even though some suggestions (e.g., providing solutions to identified problems as part of the output) were determined to be beyond the scope of the current development project, the reviewers provided useful direction for modifications of the proposed expert system.

Appraisal

The appraisal plan (see Appendix H) included criteria for evaluating the activities of the product design phase. Specifically, reviewers' responses on the product design phase instruments were compared with responses written by the developer. The agreement between reviewers' responses and the developer's responses exceeded the criterion of at least 80 percent agreement on each of the instruments. Percentages of agreement computed for the Clarity Evaluation Instrument, Internal Consistency Worksheet, Regulations

Checklist, and Interview Form were 92, 90, 90, and 92 percent, respectively.

Product Prototype

The process of developing a prototype of the product included the following components: (a) an appraisal plan identifying the major activities for this phase, (b) the development of an initial prototype, (c) prototype testing and revision, and (d) an appraisal of the activities completed during the prototype phase, including preliminary assessments of system validity and reliability. Materials required for the activities of this phase included broad-based test cases, documentation for operating the expert system prototype, and forms for collecting data regarding the consultations using the test cases. In addition, a product prototype questionnaire was developed and implemented to collect the feedback from formative evaluators testing the prototype. The various components of this phase and the respective materials and questionnaire are discussed in the following sections.

Appraisal Plan

Prototyping appears as a consistent feature of expert system development (Goodall, 1985). Using diverse cases, developers test various versions of a prototype to search for potential problems in the knowledge base and inference structure of the system (Buchanan et al., 1983; Kidd, 1984). The feedback generated by the test cases is used to improve subsequent versions of the prototype until the output is judged valid and reliable (Hofmeister, in press). To facilitate valid and reliable output of the product prototype the author designed an appraisal plan identifying the development and evaluation activities of this phase. These planned activities appear in Appendix S.

Prototype Development

At this point in the development process, the definition phase and design phase results were transferred into computer program code and documentation. This resulted in a prototype version of the product. In order to write the program code, the developer subdivided the IEP development process into components; for example, one component addressed the appropriate team members for an IEP team meeting, while another component dealt with acquiring appropriate team member signatures for approval of the IEP. Program code was written for each component and initially tested to assess the accuracy of the inference structure. Then the program components were combined until all components were linked into a single prototype.

With a completed version of the prototype, the developer began systematic testing of the product using a diverse set of test cases created by special education graduate students and the developer. The set of test cases were intentionally diverse in order to tap all aspects of the knowledge base of the system. The process of running test cases and revising the program code continued until the developer determined the prototype to be a working system for all areas of the knowledge base. At this point, the system was ready for the critique of formative evaluators.

In addition to the prototype, documentation was written to accompany the computer program. The documentation contained a description of the expert system, directions for its operation, and additional information essential to potential users. A copy of the "Documentation for the Operation of Mandate Consultant" appears in Appendix T.

Prototype Testing and Revision

At this point, the formative evaluation of the prototype extended to a group of professionals in the field of special education from the state education agency, local education agencies, and higher education. Using multiple combinations of diverse test cases, created by the developer from actual cumulative student files, reviewers of the prototype read the documentation, reviewed the test case cumulative file data, and ran consultations on the test cases using the prototype.

The reviewers completed a Test Case Data Collection Form (see Appendix U) for each completed consultation. The data form included: (a) discrepancies between the test case cumulative file data and appropriate procedures, as stated in federal and state regulations and determined by the reviewer (b) differences between the reviewer's conclusions and the conclusions generated by the consultation using the prototype, and (c) other relevant comments from the reviewer about the operation of the prototype. The reviewer's responses provided the descriptive data necessary for the cyclic process of testing and revising subsequent versions of the prototype.

Product Prototype Questionnaire

In addition to the descriptive data collected for each test case, summary data were collected using a formal questionnaire (see Appendix V). The same reviewers completed the questionnaire after finishing test case consultations. The questionnaire addressed two aspects of the prototype: (a) accuracy, and (b) operation. Accuracy issues included the clarity of questions, responses, and explanations appearing on the computer monitor during consultations, as well as the reviewer's agreement with the system's conclusions. Operational issues included the reviewer's judgment of the organization, clarity, and comprehension of the documentation, as well as the reviewer's opinions about the feasibility of school administrators actually implementing the prototype.

As before, the author used the technique described by Martuza (1977) to estimate the instrument's validity and reliability. Two content specialists reviewed the instrument and, using a four-point rating scale, judged the relevance of each item to its respective objective. The specialists' responses were used to compute a measure of interrater agreement and an index of content validity. The outcome of the data analysis supported the instrument as validly and reliably meeting its objectives. The raw data and the results of the data analysis appear in Appendix W.

Results of the Product Prototype Questionnaire

The evaluators reported the questions, responses, explanations, and conclusions generated by the prototype during the test case consultations as generally clear. The evaluators also judged the system's conclusions as agreeing with their interpretation of federal and state regulations governing IEP development. In addition, the evaluators reported that the documentation was logically organized, clear, and sufficiently comprehensive for successful operation of the prototype. Two of the evaluators noted that the prototype's rate of responding was too slow and may cause frustration for potential users.

The evaluators stated varied opinions regarding implementation of the prototype into school settings. Some believed the computer hardware required to run the system would prevent schools from acquiring it; other evaluators disagreed. Some evaluators stated that locating the required computer hardware and software in the school's administrative office would prevent many potential users from accessing the system. Other evaluators did not agree. Finally, most evaluators viewed the consultation time as feasible for successful implementation in the schools.

Appraisal

The evaluation of the product prototype phase included several components: (a) reviewing the test case results as reported on the Test Case Data Collection Forms (see Appendix U), (b) assessing the reliability of the conclusions generated by the expert system, both interuser and intrauser reliability, (c) reassessing content validity of the knowledge base, and (d) reviewing the data collected with the Product Prototype Questionnaire.

Review of Test Case Data

As described in the "Prototype Testing and Revision" section of this manuscript, numerous evaluators reviewed test case cumulative files (for example, parent permission for placement form, IEP, log of contacts with parents) and noted discrepancies between the cumulative file documentation and the procedures governing IEP development, as stated in federal and state regulations. Then evaluators ran consultations of the test cases using the prototype.

The evaluators compared the conclusions generated from the consultation with the discrepancies they noted prior to using the prototype. Interpretation of the data gathered from this process, as reported on the Test Case Data Collection Form (see Appendix U), indicated the need for specific modifications of the prototype. The cyclic process of modifying the prototype and running test cases continued until no substantive modifications were suggested by the data. That is, the agreement between the evaluator's own conclusions on test cases with the conclusions generated when the evaluator used the prototype exceeded the 80 percent criterion stated on the appraisal plan for the prototype phase.

A final component of the prototype phase included a preliminary assessment of the accuracy of the prototype's conclusions when field-representative test cases were run by the evaluators. In order to determine a standard, the developer generated conclusions from the prototype for a group of six test cases. Then the number of conclusions from the cases run by evaluators agreeing with the number of the developer's conclusions were correlated using a Pearson product moment correlation (Ferguson, 1981). The correlation described the association between the number of prototype conclusions identified when the developer and each evaluator conducted the consultations. The correlations ranged from .87 to 1.00 with a mean of .91. Correlations for the five evaluators included .87, 1.00, .87, .87, and .94, respectively. Recognizing that some error is associated with any measure, correlations of this magnitude generally indicate a relatively high degree of association (Ferguson, 1981).

Preliminary Assessment of Reliability

Establishing reliability among various users of a product, as well as by the same user over time, is an essential step to systematic product development (Borg & Gall, 1983). Therefore, the developer collected data to assess both interuser and intrauser reliability. Multiple evaluators running consultations using the same test case data provided interuser reliability data. The conclusions generated for the multiple consultations of each case were compared. Percentages of agreement among the conclusions from the multiple consultations on the same cases ranged from 71 percent to 89 percent with a mean of 83 percent. Recognizing that some error is associated with any measure, this level of agreement indicated a substantially high probability that various users of the prototype could obtain similar results (Borg & Gall, 1983).

In order to estimate intrauser reliability, the same evaluator ran the same test case at different times. The percentages of agreement for the conclusions from these test cases ranged from 67 percent to 100 percent, with a mean of

88 percent. Again, recognizing that some error is associated with any measure, this level of agreement indicated a relatively high probability that users could obtain very similar results on the same cases over time (Borg & Gall, 1983).

As part of the preliminary reliability assessment, the developer reviewed printed records of the test case consultations run by formative evaluators for each case, to identify any particular questions or responses that appeared consistently troublesome for the evaluators. Subsequent prototype modifications included clarification of these problem questions and responses.

Reassessment of Content Validity by Specialist

Several months of prototype testing and revision resulted in substantial modifications to the expert system's knowledge base rules. Therefore, a state education agency official reevaluated the content of the knowledge base and reassessed its consistency with the procedures stated in the federal and state regulations. The Regulations Checklist (see Appendix N) provided the structure for the reassessment. The state agency official reaffirmed the content of the knowledge base as congruent with the regulations.

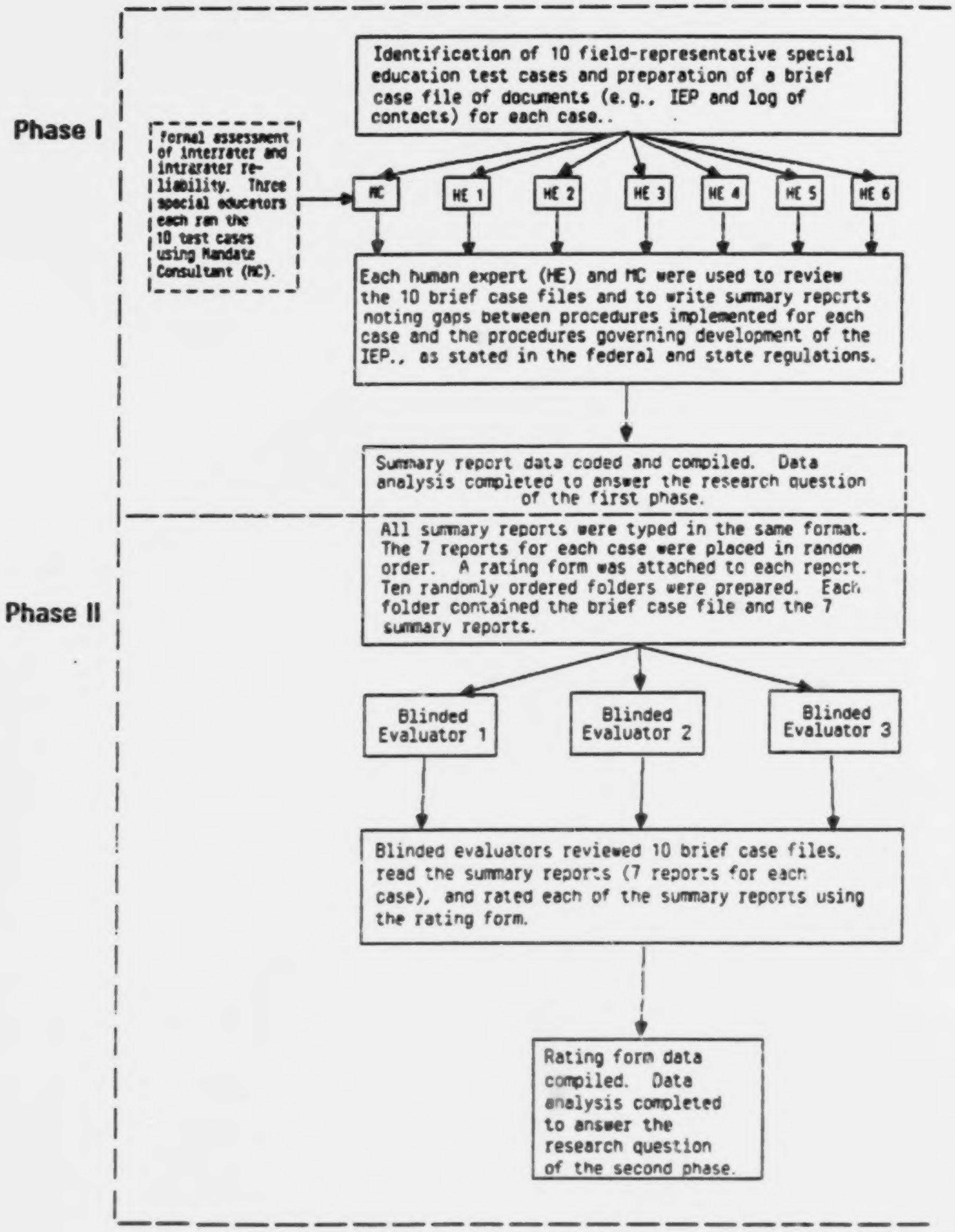
Results of the Product Prototype Questionnaire

The appraisal plan (see Appendix S) included a criterion for evaluating the results of the Product Prototype Questionnaire. The responses of evaluators completing the questionnaire were compared with responses written by the developer. The agreement between the evaluators' responses, as reported in the "Product Prototype Questionnaire" section of this manuscript, and the developer's responses exceeded the criterion of at least 80 percent agreement.

Product Validation

Design

During this activity the focus shifted from product improvement to formal assessment of the accuracy of the expert system's output. The experimental design for this procedure involved two formal evaluation phases, as depicted in Figure 2. In the first phase, six human experts reviewed the data of ten field-representative cumulative case files and provided conclusions regarding failures to implement state and federal regulatory procedures for IEP development. In addition, conclusions about the IEP procedures were generated from these same ten cases by the expert system. In the second phase, three additional human experts reviewed all the conclusions and judged their acceptability using a



Identification of 10 field-representative special education test cases and preparation of a brief case file of documents (e.g., IEP and log of contacts) for each case..

Formal assessment of interrater and intrarater reliability. Three special educators each ran the 10 test cases using Mandate Consultant (MC).

Each human expert (HE) and MC were used to review the 10 brief case files and to write summary reports noting gaps between procedures implemented for each case and the procedures governing development of the IEP., as stated in the federal and state regulations.

Summary report data coded and compiled. Data analysis completed to answer the research question of the first phase.

All summary reports were typed in the same format. The 7 reports for each case were placed in random order. A rating form was attached to each report. Ten randomly ordered folders were prepared. Each folder contained the brief case file and the 7 summary reports.

Blinded evaluators reviewed 10 brief case files, read the summary reports (7 reports for each case), and rated each of the summary reports using the rating form.

Rating form data compiled. Data analysis completed to answer the research question of the second phase.

Figure 2. Design for Product Validation

rating scale. These reviewers were "double-blinded," that is, they did not know the source of the conclusions, including not knowing that one of the sources was a computer program. The evaluators' ratings served as a basis for comparing the human expert and expert system conclusions.

This type of blinded evaluation of expert system knowledge base performance was implemented in the medical field to evaluate the expert systems, MYCIN and ONCOCIN (Yu, Fagan, et al., 1979; Hickam, et al., in press). Their two-phase design evolved from earlier evaluations of MYCIN. Their evaluation compared experts' decisions, where the answer is not clearly a "right" or "wrong," with the expert system's conclusions (Yu, Fagan, et al., 1979). Because a model for evaluating educational expert system knowledge bases had not been developed, the design used successfully to evaluate MYCIN and ONCOCIN in the field of medicine was selected for this study.

In addition, Phase I included formal assessment of interuser and intrauser reliability. This subcomponent also appears in Figure 2. Specifically, three special educators independently used the expert system to run consultations of the ten test cases. If any discrepancies in the conclusions generated by the system resulted, the developer used the conclusions obtained by two of the three special educators. The conclusions for the test cases, obtained by the special educators using the expert system, were incorporated as part of the data for the Phase I data analysis. These conclusions were also used in conjunction with the six human experts' conclusions for the "double-blinded" evaluation of Phase II.

Sample

A total of nine human experts participated in the two phases. These experts were selected from a pool of special education administrators and other leaders (for example, advocates, attorneys, and university-affiliated staff) actively involved in special education in Utah. A list of special education administrators and other leaders was obtained from the staff of the state education agency of Utah. In addition, staff of the state agency identified those leaders and administrators on the list who, in the staff's judgment, were the most qualified "experts" in the field. As the state agency staff identified the administrators, they were instructed to consider such factors as the amount of experience, diversity of experiences, and specialized training received. The state agency staff were also asked to identify other leaders in special education who were actively involved in special education issues, and who were viewed by staff of the state agency as "experts" in the field. The two groups of experts, that is, administrators and other leaders identified as most "qualified," formed the accessible population. The nine most "qualified" human experts were selected for the study. Of the sample, the three experts considered most "qualified," as based on the criteria described earlier, were selected for Phase II of the evaluation.

Six human experts were involved in the first phase of the evaluation. Their qualifications as experts are summarized in Table 2.

Three human experts acted as evaluators for the "double-blinded" phase of the study. Their qualifications appear in Table 3.

Measurement and Instrumentation

Selection and Preparation of Test Cases from the Field

To accomplish the product validation study, it was necessary to gather a set of representative special education test cases from a local education agency. Special education administrators from the local education agency (Cache County School District) identified ten special educators who were representative of the service delivery continuum provided locally. Two special educators provided services primarily on an itinerant basis, five provided services primarily in a resource room setting, and three taught in a self-contained classroom setting. Also, the ten educators provided services to students representing a myriad of handicapping conditions. The special educators

Table 2

Phase I: Human Experts' Qualifications

	Current professional position	Years of professional experience	Participated as trainer or trainee in sessions regarding implementation of regulations governing development of IEPs	Qualified as a Due Process Hearing Officer for PL 94-142
*HE 1	**LEA Director of Spec. Ed	7	yes	no
HE 2	LEA Director of Spec. Ed	12	yes	yes
HE 3	LEA Director of Spec. Ed	15	yes	no
HE 4	LEA Director of Spec. Ed	17	yes	yes
HE 5	LEA Director of Spec. Ed	22	yes	no
HE 6	LEA Director of Spec. Ed	19	yes	no
		(Mean = 15.3)		

*Human Expert

**Local Education Agency

Table 3

Phase II: Evaluators' Qualifications

Evaluator	Current professional position	Years of professional experience	Participated as trainer or trainee in sessions regarding implementation of regulations governing development of IEPs	Qualified as a Due Process Hearing Officer for PL 94-142
1	Private Consultant/ Advocate for Special Ed.	34	yes	yes
2	*LEA Administrator	25	yes	yes
3	LEA Director of Spec. Ed	15	yes	yes
		(Mean = 24.7)		

*Local Education Agency

were asked to randomly select one student from their respective student population and photocopy selected documents from the student's cumulative file. They were also asked to remove any personally identifiable information from those documents. The memorandum given to educators specifiying case selection and preparation of the cumulative case file documents appears in Appendix X.

In summary, the test case data gathered from the Cache County School District included 7 male and 3 female special education students ranging in age from five to seventeen years. The handicapping conditions presented by the cases ranged from mild specific learning disabilities (SLD) to severely intellectually handicapped (SIH). The students received a myriad of special education and related services on an itinerant basis, in a resource room setting, or in a self-contained classroom setting.

The developer prefaced each of the ten cumulative student files with a cover sheet that identified the case with a fictitious name, stated the date of the most recent comprehensive evaluation of the case, and listed the student's handicapping condition. The typical cumulative file included a parent permission for placement form, an IEP, and a log of contacts with the parent. In some instances, other documents, such as a copy of the letter inviting the parent to the IEP team meeting, were attached. Each brief cumulative file was checked to reaffirm that all personally identifiable data had been removed. Finally, for dissemination purposes, multiple copies of each file were made.

Phase I

As described earlier in the evaluation design, six human experts independently reviewed the ten case files selected from the field. To facilitate this activity, the developer mailed each human expert a packet of materials that included (a) ten brief cumulative files (b) a letter containing directions for reviewing the ten cumulative files (see Appendix Y), (c) ten cumulative file report forms (see Appendix Z), (d) an informed consent form (see Appendix AA), (e) a general information form (see Appendix BB), and (f) a suggestion notice attached to the files (see Appendix CC). The human experts read the files and then noted discrepancies between the procedures implemented for the case (as evidenced in the file documentation) and the procedures governing IEP develpment as stated in federal and state regulations.

Directions for reviewing the cumulative files. The letter (see Appendix Y) to the human experts described the specific task to be completed. The directions included possible questions that the human experts might pose as they read and reviewed the cases. The developer's intent in providing possible questions was to help the human experts structure their thinking in approaching the task. However, the developer refrained from providing specific structure such as a checklist. Providing extensive structure could have interfered with collecting valid data for answering the first research question, that is, "to what degree do expert system-generated conclusions match human expert conclusions?"

Directions for completing the cumulative file report form. The packet mailed to the human experts included ten cumulative file report forms (see Appendix Z) for recording the results of each case reviewed by the human expert. This form contained two sections: (a) summary statements, and (b) related comments and/or specific citations. Once again, the developer provided some structure for the human expert by stating some possible summary statements and related comments. But as before, the developer avoided providing extensive structure that may have interfered with collecting valid data for the purpose of this evaluation.

Use of "suggestion" notice. Recognizing that the directions contained in the letter were quite lengthy, the developer provided another prompt to the human experts in the form of a "suggestion" (see Appendix CC). The "suggestion" was attached to the front of the case files to focus the human experts' attention on the desired task. After reviewing each case file and completing the ten summary report forms, the human experts were directed to return the completed packet to the developer.

Data collection for the consultations using Mandate Consultant. Simultaneous to the activity with the human experts, three special educators were independently completing consultations with the expert system using the documentation contained in the ten brief cumulative files. The author selected to use three special educators so that if any of the conclusions were discrepant, the conclusions generated by two of the three consultations would be used as the data generated by the expert system. Using multiple special educators for running the consultations also provided the data for a formal assessment of the interuser and intrauser reliability. The results of the consultations using the expert system were compiled to reflect the conclusions of the majority and transfered to cumulative file report forms like those completed by the human experts.

Coding of the cumulative file report data. At this point, a critical step in the study involved the coding and compilation of the data contained in the summary reports from the experts. This was done by determining when the conclusions of an expert matched or did not match the conclusions of another expert. Recognizing the potential experimenter bias, a special education graduate student, who was unfamiliar with the system, completed the coding. First, the coder read the cumulative file reports and eliminated any conclusions considered irrelevant for the purpose of this study, based on the following criteria: (a) The conclusions regarded issues not addressed in the federal or state regulations (some of these items may have some value, but they are considered "good practices" rather than a requirement); and (b) the conclusions regarded issues addressed in the federal or state regulations, but the issues were not the focus of this study (this project focused on the regulations specific to the IEP team process resulting in the development of an IEP).

With the extraneous conclusions eliminated, the coder categorized the remaining conclusions as agreements or disagreements using the following categories:

1. Agreement. This category included conclusions stated by both human experts and the expert system. The language may have differed (a summary statement or a related comment/citation), but the intent or substance were judged to be the same.
2. Disagreement. This category included:
   (a) conclusions stated by one source (human expert or expert system) that were contradicted by statements from another source (human expert or expert system), and/or
   (b) conclusions stated by one source (human expert or expert system) that were not stated

by another source (human expert or expert system).

This categorized data provided the data for analysis in responding to the research question of Phase I.

Phase II

Experimental validity precautions. To control for threats to experimental validity, the developer took several precautions in preparing the cumulative file reports and accompanying materials for the "double-blinded" evaluation. In order to meet the definition of a "double-blinded" study, it was essential that the evaluators of this phase not know the source of any of the cumulative file reports; furthermore, the evaluators were not to know that one of the sources was an expert system. The precautions taken included the following.

First, all cumulative file reports were typed in a similar format and printed on either dot matrix or letter-quality printers. The printed originals were photocopied to produce the required copies for the study. Second, to eliminate any order pattern, the seven reports (six from human experts and one from the expert system) for each of the ten cases were placed in random order. The seven randomly-ordered reports were inserted into pocket folders prepared for each case. Finally, the ten pocket folders were placed in a differing random order for each evaluator's packet.

Directions for completing rating task. As depicted in Figure 2, three human expert evaluators independently reviewed and rated the cumulative file summary reports. The evaluators read the same ten cumulative files as the human experts of Phase I. Then the evaluators reviewed the cumulative file reports written by the experts and rated each report according to its acceptability, based on a four-category rating scale. To facilitate this activity, the developer mailed each evaluator a packet of materials that included: (a) a letter containing directions for reviewing the cumulative files, the reports, and for rating the ten cumulative reports (see Appendix DD); (b) an informed consent form (see Appendix EE); (c) a general information form (see Appendix FF); (d) ten pocket folders, each containing a brief cumulative file, and seven cumulative file reports with attached rating forms (see Appendix GG); and (e) a copy of the directions and suggestion notice provided to the human experts of Phase I. The letter (see Appendix DD) to the evaluators described the specific task to be completed. The evaluators were instructed to (a) review each cumulative file, (b) formulate their own judgments about the case, and (c) complete the rating forms attached to each of the seven cumulative file reports.

The Rating Form. The primary data collection instrument for Phase II was the Rating Form (see Appendix GG). The four-category rating scale was modeled after the one validated as part of the ONCOCIN evaluation. (Hickam, et al., in press). Researchers from the ONCOCIN project conducted a pilot study, using the four-category rating scale, and found the scale to appropriately represent experts' opinions. The scale implemented for this study included: (a) 1 - Ideal: The information summarized in the report is synonymous with what I would have written, (b) 2 - Acceptable: The information summarized in the report differs from what I would have written, but it is acceptable, (c) 3 - Less than acceptable: The information summarized in the report is inaccurate and/or inadequate; however, I would consider these deficiencies minor, and (d) 4 - Unacceptable: The information summarized in the report is inaccurate and/or inadequate, and I would consider these deficiencies major.

The evaluators completed a rating form containing the four-category rating scale for each cumulative file report. If evaluators rated a report as less than acceptable or unacceptable, they identified specific deficiencies of the report. This descriptive information was reviewed to determine if particular deficiencies were consistently reappearing in the cumulative file reports.

After completing the rating task, the evaluators returned the completed packet to the developer. The information from the completed rating forms provided the

Table 5

Number of Interexpert Agreements (A) and Disagreements (D) for Mandate Consultant (MC) and Six Human Experts (HE)

	MC		HE 1		HE 2		HE 3		HE 4		HE 5		HE 6		Mean for each case	
CASES	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D
Alan	9	-	4	-	9	-	4	-	4	-	4	-	8	-	6.0	-
Brad	4	-	0	-	4	-	3	-	1	-	2	-	2	-	2.3	-
Craig	17	-	5	-	9	-	12	-	9	-	5	-	5	-	8.9	-
Joe	14	-	3	-	12	-	11	-	9	-	13	-	12	-	10.6	-
Margaret	6	2	-	-	6	2	6	-	3	-	3		-	-	3.4	.6
Maryann	4	1	-	-	2	1	2	-	-	-	3	-	3	-	2.0	.3
Richard	12	-	3	-	9	-	10	-	7	-	7	-	-	-	6.9	-
Robert	1	-	-	-	3	-	3	-	-	-	1	-	2	-	1.4	-
Russell	7	1	-	-	1	1	5	1	-	-	-	-	4	1	2.4	.6
Sherri	2	-	1	-	2	-	3	-	-	-	-	-	1	-	1.3	-
*Total	76	4	16	-	57	4	59	1	33	-	38	-	37	1	45.1	1.5
**Avrg.	7.6	.4	1.6	-	5.7	.4	5.9	.1	3.3	-	3.8	-	3.7	.1		

*Total number of interexpert agreements and disagreements by expert.

**Average number of interexpert agreements & disagreements per case.

data for analysis in response to the research question of Phase II.

Data Analysis

Evaluations of knowledge-based expert systems create difficulty because there can be many differences of opinion, even among experts (Shortliffe, 1981). Thus, the selection of a "gold standard" by which to measure a system's performance is also difficult. Gaschnig et al., (1983) suggest a viable alternative. They recommend presenting the same information to human experts and the expert system, then comparing expert system output to what the human experts state as the correct answer.

In order for Mandate Consultant to be judged accurate, it needed to produce conclusions about implementation of regulatory procedures for the development of an IEP which were similar to human experts' conclusions. Thus, data were collected regarding conclusions produced by the expert system and human experts, and these data were analyzed and compared in terms of similarities and differences between and among the conclusions. The two major research questions used to guide the analysis of data were initially stated in the "Introduction."

Phase I

The research question of Phase I asked: To what degree do expert system-generated conclusions match human expert conclusions in terms of noting discrepancies with special education regulatory procedures for the development of IEPs?

The data from the cumulative file reports produced by the experts and coded by the special education graduate student were analyzed and tabulated to determine: (a) the total number of conclusions produced by the experts, (b) the total number of interexpert agreements and disagreements for the conclusions, (c) the associations between the number of interexpert agreements reported for the cases (using the Pearson product-moment correlation coefficient), (d) the total number of conclusions most frequently noted by the experts, (e) rankings of the experts based on the previous totals, and (f) the association between the sets of rankings of experts based on the previously reported totals (using Kendall's coefficient of rank correlation).

The tabulation of total conclusions, interexpert agreements and disagreements, and most frequently-noted conclusions provided the data for computing the Pearson product-moment correlation coeffecient (r) and Kendall's coefficient of rank correlation (Tau). The Pearson r quantified the association between the number of agreements for each expert with every other expert. Kendall's Tau provided a measure of the association between three sets of expert rankings based on the number of conclusions, the number of interexpert agreements, and the number of most frequently-noted conclusions. Because the sets of ranks were based on related data, a substantial association was expected. Nonetheless, Kendall's Tau provided a technique for quantifying these associations.

These comparisons were used to evaluate the degree to which the experts' conclusions matched in terms of implementing regulations governing the development of IEPs. When all tabulations, computations, and analyses were complete, the Phase I research question was answered.

Phase II

The research question of Phase II asked: To what degree are expert system-generated conclusions or human expert conclusions rated as acceptable (based on a rating scale of 1-ideal, 2-acceptable, 3-less than acceptable, and 4-unacceptable) by a panel of human expert evaluators?

During Phase II of the formal evaluations, human expert and expert system-generated conclusions stated on case reports were independently compared by a panel of three human expert evaluators. The rating data from this activity were compiled to determine: (a) a measure of the interevaluator reliability, (b) sets of rankings of the experts based on the ratings by the three evaluators, (c) the association between the sets of rankings (using Kendall's coefficient of concordance), and (d) percentage of expert case reports judged acceptable or unacceptable based on the evaluators' ratings.

The expert rankings resulting from each evaluator's ratings provided the data for computing Kendall's coefficient of concordance (W). Kendall's W quantified the association between the three sets of rankings of the experts.

In addition, the number of cases (percentage) in which the reports of the expert system or human were rated as at least "acceptable" (that is, 1-ideal or 2-acceptable) by one evaluator were calculated. With ten case summaries and three evaluators, thirty ratings were generated for each expert. Thus, fractions with the number of ratings of at least "acceptable" in the numerator and the total number of ratings per expert (i.e., 30) in the denominator were converted to percentages. The percentage obtained for each expert provided a basis for comparing human and expert system-generated conclusions.

Furthermore, the number of cases (percentage) in which the conclusion of the expert system or human were rated as "acceptable" by the majority of the evaluators was calculated. Thus, a fraction with the number of ratings as "acceptable" by the majority of evaluators in the numerator, and the total number of cases (i.e., ten) in the denominator were converted to percentages. The percentage obtained for each expert provided a basis for comparing human and system-generated conclusions.

Finally, each rating form that indicated the conclusions were less than acceptable or unacceptable included comments from the evaluator stating the specific deficiencies of the conclusions. The deficiencies were reviewed and analyzed to identify particular weaknesses in the conclusions from both the human experts and the expert system.

When the calculations and resulting percentages for each expert, the computation of the coefficient of concordance, and the analysis of the deficiencies noted by the evaluators were completed, the Phase II research question was answered. The two percentages provided different perspectives (that is, "acceptable" by one evaluator or "acceptable" by the majority of evaluators) for answering the same question. The analysis for Phase II complemented the analysis for Phase I by providing an additional level of comparison of agreement among and between experts.

Assessment of reliability. As a part of the validation study, the reliability of consultation outcome between and within users of Mandate Consultant was formally assessed. Percentages of agreement for (a) the conclusions, and (b) all responses generated by the expert system were computed. Data resulted from three special educators independently running the same cases using the expert system, as well as three special educators each running the same cases at different times. The percentages of agreement provided measures of both interuser and intrauser reliability, respectively.

RESULTS AND DISCUSSION

Results

A summative evaluation of the expert system, Mandate Consultant (MC), was conducted in two phases involving nine human experts. The purpose of the evaluation was to assess whether the knowledge-based expert system accurately emulated the knowledge of experienced special educators regarding regulatory procedures for developing IEPs. Phase I of the evaluation involved human expert and expert system review of ten representative cumulative student files. Phase II involved judgments of the reviewers' conclusions by human evaluators who were "blinded" regarding the source of the conclusions, including not knowing that one source was an expert system.

The specific research questions posed by the author appear in the "Introduction" chapter. A presentation of the results, as they relate to each question, follows. In addition, the author concludes this section with the results of the formal assessment of user reliability.

Phase I

The research question of Phase I asked: To what degree do expert system-generated conclusions match human expert conclusions in terms of noting discrepancies with special education regulatory procedures for the development of IEPs?

The conclusions produced by the experts, both human and expert system, following their independent review of the ten cumulative student files, provided the basic units of measurement for Phase I. The conclusions, as coded by the special education graduate student, appear in Appendix HH.

Number of Conclusions Generated

The coded data were initially tabulated based on the number of conclusions produced by the experts for each of the ten cases. This information appears in Table 4. The cells in the body of the table indicate the number of conclusions reported by one expert for a particular cumulative student file. These files are referred to by fictitious case names. The table shows that Mandate Consultant produced more conclusions than any of the human experts. Also, on the average, four of the human experts produced fewer than two conclusions for each case.

Number of Interexpert Agreements and Disagreements

Next the coded data were tabulated to indicate the number of times the conclusions of one expert agreed or disagreed with any other expert. This number of interexpert agreements and disagreements are reported in Table 5. The individual cells in the table reflect the number of agreements and disagreements of one expert for a particular

Table 4

Number of Mandate Consultant (MC) and Human Expert (HE) Conclusions

Cases	MC	HE 1	HE 2	HE 3	HE 4	HE 5	HE 6	Average # of conclusions by case
Alan	6	1	3	1	1	1	2	2.0
Brad	4	0	3	4	2	1	1	2.1
Craig	8	3	4	5	4	3	3	4.3
Joe	7	3	3	4	2	4	4	3.9
Margaret	6	1	4	3	1	1	0	2.3
Maryann	7	1	3	2	0	1	1	2.1
Richard	6	2	6	4	2	2	1	3.3
Robert	4	0	3	3	0	2	3	2.1
Russell	6	2	2	5	0	0	3	2.6
Sherri	2	1	1	4	0	0	1	1.3
*Total	56	14	32	35	12	15	19	26
**Average	5.6	1.4	3.2	3.5	1.2	1.5	1.9	2.6

* Total number of conclusions reported by experts.

** Average number of conclusions per case.

case. Notably, Mandate Consultant achieved more agreements than the human experts. Also, Mandate Consultant and Human Expert 2 produced the majority of disagreements among the experts. While these disagreements were analyzed, they were few in number (approximately 3% of all conclusions) and did not substantially influence the interpretation of the results.

The number of agreements reported for each expert by case in Table 5 provided the data for computing Pearsor product moment correlations (r) between each pair of experts. A matrix showing each correlation appears in Table 6. As one might expect, the correlations are relatively strong, indicating that the number of agreements for experts tend to increase or decrease together. The strongest correlations accompany Human Expert 4, Mandate Consultant, and Human Expert 2.

Most Frequently-Noted Conclusions

A final tabulation concerned those conclusions most frequently noted by experts for each case. The average number of conclusions produced by any expert for a case approached three (specifically, 2.6). Thus, the three conclusions noted most often by experts for a particular case were counted in this tabulation. The number of most frequently-noted conclusions are displayed in Table 7. The cells in the body of the table reflect the number of most

Table 6

Correlation Matrix: Pearson r's for Each Pair of Experts Based on Number of Interexpert Agreements

	MC	HE 1	HE 2	HE 3	HE 4	HE 5	HE 6	Mean correlation of the number interexpert agreements
*MC		.85	.82	.93	.94	.73	.53	.80
**HE 1	.85		.81	.73	.85	.58	.52	.72
HE 2	.82	.81		.79	.94	.88	.58	.80
HE 3	.93	.73	.79		.94	.75	.37	.75
HE 4	.94	.85	.94	.94		.85	.50	.84
HE 5	.73	.58	.88	.75	.85		.65	.74
HE 6	.53	.52	.58	.37	.50	.65		.53

*Mandate Consultant

**Human Expert

Table 7

Number of Most Frequently-Noted Conclusions By Mandate Consultant (MC) and Human Experts (HE)

CASES	MC	HE 1	HE 2	HE 3	HE 4	HE 5	HE 6
Alan	3	1	3	1	1	1	2
Brad *	2	0	3	1.5*	.5*	1	1
Craig**	2	0	2	2	2	0	1
Joe	3	0	3	2	2	3	2
Margaret**	2	0	2	2	1	1	0
Maryann	2	0	2	2	0	1	1
Richard	3	1	2	3	2	2	0
Robert***	0	0	1	1	0	1	0
Russell**	2	0	1	1	0	0	2
Sherri	1	1	1	3	0	0	1
Total of most common conclusion	20	3	20	18.5	8.5	10	10

* A tie existed for the conclusion appearing as the 3rd most frequently-noted, so the tied conclusions were each assigned a value of .5.

** These cases had only two frequently-noted conclusions.

*** This case had only one frequently-noted conclusion.

frequently-noted conclusions reported by one expert for a particular case. The maximum number of frequently-noted conclusions for any particular case is three, unless otherwise indicated. As seen from the table, Mandate Consultant, Human Expert 2, and Human Expert 3 produced the greatest number of most frequently-noted conclusions.

To assist in summarizing the tabulated information in Tables 4, 5, and 7, the author prepared the bar graph that appears in Figure 3. The graph displays the totals reported in previous tables for visual comparison. It dramatizes the variation evident in the conclusions produced by the experts.

Kendall's Coefficient of Rank Correlation

Finally, the data regarding expert conclusions were summarized by ranking the experts based on the total number of conclusions, interexpert agreements, and most frequently-noted conclusions as reflected by Tables 4, 5, and 7 (see Table 8). The rankings provided a basis for computing correlations to describe the association between the various pairs of rankings, that is, Kendall's Coefficient of Rank Correlations (Tau).

The results of the computations appear in Table 9. Recognizing that values for Kendall's Tau range from -1 to +1, the obtained correlations indicated strong associations between the pairs of rankings (Tau = .81, .75, .75).

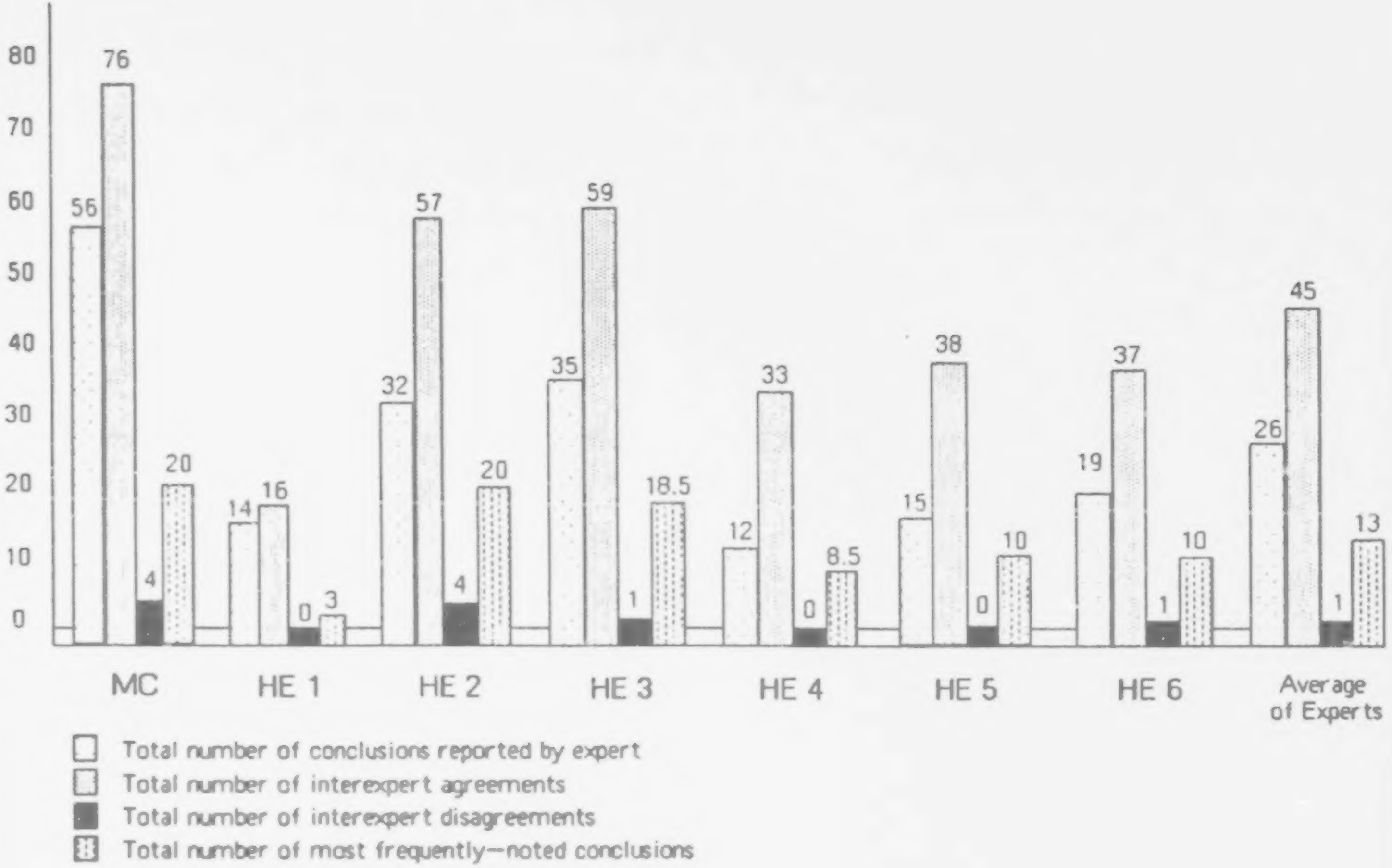


Figure 3. Totals by Expert Across the Ten Cases

Table 8

Rankings of Experts by Total Conclusions, Interexpert Agreements, and Most Frequently-Noted Conclusions

Ranking based on total number of conclusions reported by Expert			Ranking based on total number of interexpert agreements			Ranking based on total number of most frequently-noted conclusions reported by Expert		
Rank	Expert	# of Conclusions	Rank	Expert	# of Agreements	Rank	Expert	# of Common Conclusions
1	*MC	56	1	MC	76	1	MC	20
2	**HE 3	35	2	HE 3	59	1	HE 2	20
3	HE 2	32	3	HE 2	57	3	HE 3	18.5
4	HE 6	18	4	HE 5	38	4	HE 5	10
5	HE 5	15	5	HE 6	37	4	HE 6	10
6	HE 1	14	6	HE 4	33	6	HE 4	8.5
7	HE 4	12	7	HE 1	16	7	HE 1	3

*Mandate Consultant

**Human Expert

Table 9

Results of Kendall's Coefficient of Rank Correlation (Tau)

	Measure of Disarray(S)	Kendall's Tau
Correlation of rankings for total number of conclusions and total number of inter-expert agreements by Expert	S = 17*	Tau = .81
Correlation of rankings for total number of the inter-expert agreements and total number of most frequently-noted conclusions by Expert	S = 15**	Tau = .75
Correlation of rankings for total number of conclusions and total number of most frequently-noted conclusions by Expert	S = 15**	Tau = .75

* For N = 7, the probability associated with an observed value of S = 17 is .01 (Ferguson, 1981).

** For N = 7, the probability associated with an observed value of S = 15 is .03 (Ferguson, 1981).

Additionally, statistical significance was determined by testing the significance of the measures of disarray (S) for each pair of ranks.

Since rankings based on related data are expected to show association, these correlations and measures of disarray are intended to supplement and summarize the other data analysis.

Phase II

The research question of Phase II asked: To what degree are expert system-generated conclusions or human expert conclusions rated as acceptable (based on a rating scale of 1-ideal, 2-acceptable, 3-less than acceptable, and 4-unacceptable) by a panel of human expert evaluators?

The ratings reported by the evaluators, following their independent review of the expert conclusions, both human and expert system, provided the basis for measurement in Phase II. The evaluators' independent ratings for each expert's conclusions were compiled and tabulated by the author. These tables appear in Appendix II.

Interevaluator Reliability

The compiled ratings were initially aggregated to assess the interevaluator reliability of the ratings. Because of the variability associated with rating scale instruments designed to measure complex behaviors such as reviewing cumulative student files, noting discrepancies, and writing case reports, categories 1 and 2 (ideal and acceptable) and categories 3 and 4 (less than acceptable and unacceptable) were collapsed into two categories (Martuza, 1977; Borg & Gall, 1983). The number of agreements between and among experts using the two "collapsed" categories were tallied and then divided by the total number of cases rated to obtain percentages of agreement. These computations resulted in the interevaluator reliability coefficients displayed in Table 10. Reliability coefficients were computed for each combination of evaluators, as well as across all evaluators.

Table 10

Interevaluator Reliability Coefficients for Interevaluator Ratings

	Number of Agreements / Total Number Cases Rated	Interevaluator Reliability Coefficient
Evaluators 1 & 2	44 / 70	.63
Evaluators 2 & 3	48 / 70	.69
Evaluators 1 & 3	43 / 70	.61
All evaluators	35 / 70	.50

Kendall's Coefficient of Concordance Correlation

Next the independent ratings from the evaluators were tabulated to reflect the specific rating results for each expert by evaluator. This information appears in Table 11. The cells in the body of the table indicate the number of cases for one expert which the evaluator judged to fit that specific category. The totals for the categories indicate that, in general, most of the 210 case reports containing expert conclusions were judged to be less than acceptable (that is, 94 case reports) or unacceptable (that is, 52 case reports).

The information from Table 11 provided a basis for rank ordering the experts according to the number of expert case reports judged to be ideal or acceptable by the respective evaluators. These rankings, as displayed in Table 12, provided the data for computing Kendall's coefficient of concordance correlation (W) to describe the association between the three sets of rankings of the experts.

The results of the computations appear in Table 13. The table contains the set of rankings from each evaluator with tied ranks assigned the same value. The obtained value of W indicates that relatively strong agreement exists among the three sets of rankings. In addition, a test of the statistical significance of W, based on the measure of disarray, shows the obtained value to be statistically significant at the .05 level. This implies that the

Table 11

Evaluator Ratings of Mandate Consultant (MC) and Human Expert (HE) Case Reports by Evaluator

	Evaluator 1				Evaluator 2				Evaluator 3			
Experts	Ideal	Acceptable	Less than Accept	Unacceptable	Ideal	Acceptable	Less than Accept	Unacceptable	Ideal	Acceptable	Less than Accept	Unacceptable
MC		7	3		1	5	4			4	3	3
HE 1			9	1		4	4	2			1	9
HE 2	1	7	2			6	4			4	6	
HE 3	1	3	6			3	7		1	5	3	1
HE 4		3	7				6	4			2	8
HE 5		1	7	2			7	3			1	9
HE 6		3	7			5	4	1			1	9
Totals	2	24	41	3	1	23	36	10	1	13	17	39
*Average			2.6				2.8				3.3	

*Average score for expert case reports using a rating scale of
1 = Ideal
2 = Acceptable
3 = Less than Acceptable
4 = Unacceptable

Table 12

Rank Ordering of Mandate Consultant (MC) and Human Expert (HE) Based on Percent of Reports Judged as Acceptable

Evaluator 1			Evaluator 2			Evaluator 3		
Rank	Expert	Percent	Rank	Expert	Percent	Rank	Expert	Percent
1	HE 2	80	1	MC	60	1	HE3	60
2	MC	70	1	HE 2	60	2	MC	40
3	HE 3	40	3	HE 6	50	2	HE 2	40
4	HE 4	30	4	HE 1	40	4	HE 1	0
4	HE 6	30	5	HE 3	30	4	HE 4	0
6	HE 5	10	6	HE 4	0	4	HE 5	0
7	HE 1	0	6	HE 5	0	4	HE 6	0
*Average								
	37%			34%			20%	

*Average percentage of Expert case reports judged as acceptable.

Table 13

Kendall's Coefficient of Concordance Correlation (W) Results: Three Sets of Evaluator Rankings for Seven Experts

	*MC	**HE 1	HE 2	HE 3	HE 4	HE 5	HE 6
Evaluator 1	2	7	1	3	4.5	6	4.5
Evaluator 2	1.5	4	1.5	5	6.5	6.5	3
Evaluator 3	2.5	5.5	2.5	1	5.5	5.5	5.5

Measure of Disarray (S) = 178.4***

Kendall's Coefficient of Concordance (W) = .77

*Mandate Consultant

** Human Expert

***Critical value of S (for M = 3 & N = 7) = 157.3, p < .05 (Siegel, 1956).

consistently high ranking of experts, such as Mandate Consultant or Human Expert 2, by independent 4 evaluators may be attributed to positive characteristics of the respective expert.

Percent of Expert Case Reports Judged Acceptable or Unacceptable

Another analysis of the data involved the computation and tabulation of the percent of expert case reports judged ideal or acceptable by one evaluator, ideal or acceptable by a majority of the evaluators, or unacceptable by a majority of the evaluators. These data appear in Table 14. The cells of the table indicate the percent and number of cases for one evaluator fitting the particular category. Visual inspection of the tabled information shows the inverse relationship of the percent of cases judged as acceptable (i.e., ideal or acceptable) and the percent of cases judged unacceptable. Notably, Mandate Consultant and Human Expert 2 produced the case reports with the conclusions most often judged as acceptable.

To assist in analyzing the information from Table 14, the author prepared the bar graph that appears in Figure 4. The graph displays the percentages from Table 14 for visual comparison. Figure 4 clearly shows the similar performance of Mandate Consultant and Human Expert 2 as judged by the evaluators.

Table 14

Percentage of Expert Case Reports Judged Acceptable or Unacceptable

Experts	% of cases (# of cases) in which report rated as acceptable by one evaluator (n = 30)	% of cases (# of cases) in which report rated as acceptable by a majority of evaluators (n = 10)	% of cases (# of cases) in which report rated unacceptable by a majority of evaluators (n = 10)
MC	57% (17)	60% (6)	0% (0)
HE 1	13% (4)	0% (0)	20% (2)
HE 2	60% (18)	60% (6)	0% (0)
HE 3	43% (13)	30% (3)	0% (0)
HE 4	10% (3)	0% (0)	40% (4)
HE 5	3% (1)	0% (0)	30% (3)
HE 6	27% (8)	20% (2)	10% (1)
Average	30% (9.1)	24% (2.4)	14% (1.4)

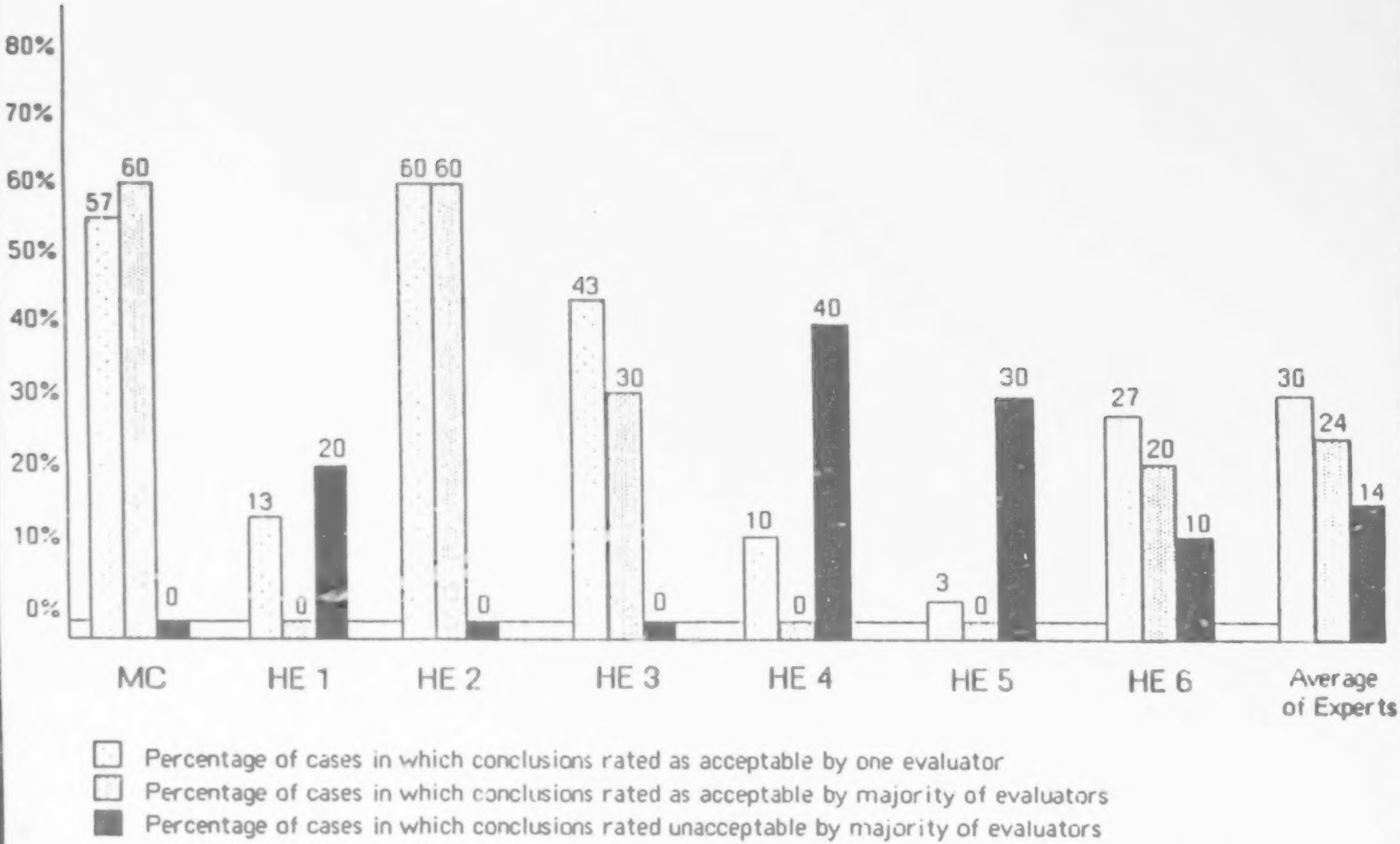


Figure 4. Percentage of Expert Case Reports Judged Acceptable

Deficiencies in Conclusions Noted by Evaluators

In addition to the quantitative data analysis, the author reviewed the specific deficiencies noted by the evaluators on rating forms indicating expert case reports as less than acceptable or unacceptable. Examples of deficiencies for the Mandate Consultant-generated reports included: (a) The evaluator questioned the conclusion that the IEP did not include a timeline for annual review; the evaluator indicated that even though no date was stated, there was no reason to believe that the local education agency would not meet the annual review requirement, (b) The evaluator questioned the conclusion that the cumulative file information did not document whether the local education agency took any actions to insure that the parent understood the proceedings at the IEP team meeting; the evaluator noted that English was reported as the primary language in the home; thus, he assumed that the parent understood the proceedings; and (c) The evaluator noted that the conclusions did not include the local education agency's failure to develop an IEP within thirty days of determining that the student required special education.

Examples of deficiencies for the human expert-generated reports included: (a) The evaluator noted that the expert did not include the lack of appropriate IEP team membership, objective criteria or evaluation procedures for reviewing student progress on goals or short-term objectives, goals for a related service, or projected dates for the initiation or anticipated length of services; (b) the evaluator noted that conclusions included opinions rather than procedures contained in federal and state regulations; and (c) the evaluator noted that some reports had so few conclusions that they provided little, if any, useful information.

In summary, those deficiencies noted in the Mandate Consultant-generated conclusions tended to be differences in the interpretation of the regulations, while the deficiencies in the human expert-generated conclusions concerned the expert's failure to note items required by the regulations.

Formal Assessment of User Reliability

During Phase I, special educators ran consultations using Mandate Consultant with the information from the ten cumulative student files. Three special educators independently completed consultations for each case which provided data for assessing interuser reliability. In addition, a sampling of the cases were rerun by the special educators at a later time, providing data for assessing intrauser reliability. The author reviewed records of the consultations and tabulated the number of agreements between users and within users for the conclusions resulting from the consultation, as well as for all the responses provided by the users during the consultations. The number of agreements for conclusions or responses was divided by the total number of conclusions or responses to produce measures of interuser and intrauser reliability. These reliability coefficients appear in Table 15. In general, reliability coefficients of this magnitude (i.e., means from .90 to .95) indicate relatively strong agreement between the different users, as well as for the same user over time.

Table 15

Reliability Measures for Phase I Users of Mandate Consultant

Reliability Coefficients

	For Conclusions		For All Responses	
	Mean	Range	Mean	Range
Interuser Reliability	.94	.67-1.00	.90	.76-1.00
Intrauser Reliability	.95	.75-1.00	.94	.90-1.00

Discussion

Phase I

In general, the results from this phase of the evaluation demonstrated that conclusions generated by the expert system match substantially with conclusions of human experts. However, the human expert behavior varied widely. In some instances, human experts comprehensively noted the failures to implement special education regulatory procedures for the development of IEPs, while in other instances, human experts noted few, if any, failures. Thus, the results of this phase indicated that the conclusions generated by Mandate Consultant generally matched the conclusions of the "better" human experts, and exceeded the conclusions of the majority of the human experts.

Mandate Consultant Produced the Most Information

The author considered several related factors in judging Mandate Consultant's performance, that is, the number of conclusions generated, the number of interexpert agreements and disagreements, and the number of most frequently-noted conclusions. Most notable about the number of conclusions generated by the experts were the few conclusions generated by the majority of human experts. The information displayed in Table 4 revealed that four of the human experts reached fewer than two conclusions per case. Clearly, no standard existed regarding the appropriate number of conclusions for cases, but the evaluators of Phase II rated numerous expert case reports as inadequate because of the few conclusions reported by the expert.

While two of the human experts (specifically, Human

Expert 2 and Human Expert 3) identified substantially more conclusions than the other human experts, none identified as many as Mandate Consultant. This finding supports the outcome of the Colbourn (1982) study. Colbourn developed and validated an expert system to assist educators in the diagnosis of reading problems. Her evaluation involved comparing expert system-generated diagnosis with human diagnosis. She found that the expert system provided more detailed information than human diagnostic reports. Such was the case with Mandate Consultant. It appeared that the extensive knowledge base contained in the structure of an expert system allowed it to generate a greater amount of specific information than many human experts typically generate.

Strong Interexpert Agreement for Mandate Consultant Conclusions

Equally important as the amount of detailed information produced by the experts was the validity of the information. In Phase I of the evaluation, some primary evidence for validity appeared in the number of agreements and disagreements for one expert with every other expert (Table 5). Again, Mandate Consultant had the greatest number of agreements, followed by Human Experts 2 and 3, respectively. The author expected the number of conclusions to be related to the number of interexpert agreements, but the significant frequency of other experts agreeing with the conclusions of Mandate Consultant strengthened the case that Mandate Consultant provided substantial amounts of valid information.

While the majority of data from Table 5 reported interexpert agreements, other data showed occasional disagreements between the conclusions of Mandate Consultant and the human experts. However, these disagreements appeared to result from different interpretations of the same cumulative file data. For example, Human Expert 2 reported that the student data for the case named "Margaret" showed that the IEP was developed within the 30-day timeline specified in the regulations, while the three special educators running the consultations using Mandate Consultant indicated that Margaret's IEP was not developed within the 30-day timeline. In another instance, Human Expert 2 noted that the amount of time "Russell" was spending in the regular education setting was noted on the IEP, while Mandate Consultant, Human Expert 3, and Human Expert 6 noted that the IEP did not specify the amount of time. Thus, the disagreements between Mandate Consultant and the human experts did not show weaknesses in the knowledge base of the expert system; rather, the disagreements appear to indicate that experts using the same cumulative file may interpret that data differently.

Most Frequently-Noted Conclusions Add to Case for Mandate Consultant

When comparing the conclusions of experts, another dimension regarded which experts were reporting the conclusions most frequently noted by the other experts. That is, were any experts noting conclusions considered irrelevant or trivial rather than the conclusions typically reported by other experts? As the data in Table 7 indicated, Mandate Consultant identified many conclusions noted by other experts. Mandate Consultant's performance compared to the "best" of the human experts. This result complemented the previous findings, showing that Mandate Consultant produced the most conclusions and had the greatest number of agreements with other experts.

The information in Table 8, Table 9, and Figure 3 quantitatively and graphically demonstrated that the conclusions of Mandate Consultant at least matched, if not exceeded, the conclusions of human experts. The strong association of various rankings of the experts, indicated by Kendall's Tau, provided summative evidence that Mandate Consultant generated conclusions considered consistently strong across three related factors as well as for each of the factors individually.

Phase II

The results of the "double-blinded" evaluation using human expert evaluators provided substantial evidence that Mandate Consultant-generated conclusions comparable to the "better" human experts, and more acceptable than the majority of human experts. A discussion of issues relevant to these results follows.

Interevaluator Reliability

The basic measurement for this phase involved the ratings of three human expert evaluators for the case reports containing expert conclusions. The credibility of interpretations based on the ratings required an assessment of the reliability of the ratings reported by the evaluators.

Reliability coefficients for the evaluators appeared in Table 10. Notably, the reliability coefficients were not particularly strong. However, when one considers the complex behavior judged by the evaluators, the coefficients appear typical. Borg and Gall (1983) emphasized the difficulty of rating complex human behavior using rating scales. They suggested that when one uses a rating scale with more than three categories (such as the instrument used in this evaluation), an appropriate standard may be "satisfactory" reliability, that is, consistent rating in light of the complexity of the behavior measured and the number of categories on the rating instrument. Recognizing that the evaluators' task in this study required judgement of behaviors involving reviewing, interpreting, and stating conclusions about cumulative file data on a four-category rating scale, the obtained reliability coefficients appeared to meet the "satisfactory" standard described by Borg and Gall (1983).

In addition, the reliability coefficients should be assessed in light of the obtained value for Kendall's coefficient of concordance (W). Kendall's W indicated a strong association between the independent rankings of the experts by the three evaluators. Even though the reliability coefficients implied that one evaluator may have rated the expert reports more or less critically than another evaluator, a consistent pattern of rankings existed across all three sets of independent rankings.

The Practical Significance of Kendall's Coefficient of Concordance

As displayed in Tables 12 and 13, Kendall's coefficient of concordance (W) was computed to describe the association between evaluators' independent rankings of the experts. Notably, the obtained value for Kendall's W was statistically significant. Equally important is the practical significance of this result. That is, consistently high rankings by independent evaluators of such experts as Mandate Consultant and Human Expert 2 may be attributed to characteristics of these experts that allowed them to generate acceptable conclusions. While the specific characteristics of experts were not studied, it appears reasonable to hypothesize that the extensive knowledge base and the structure of the expert system contributed to its performance as compared to the majority of human experts.

Evaluators Judged the Majority of Mandate Consultant's Conclusions Acceptable

The strong performance of Mandate Consultant-generated conclusions emerged from the information displayed in Table 14 and Figure 4. The quantitative and graphic information provide substantial evidence supporting the case that Mandate Consultant generated conclusions comparable to the conclusions of the "best" human expert, and more acceptable than the majority of human expert-generated conclusions.

Trends in the Deficiencies Noted by the Evaluators

Examples of deficiencies of the expert conclusions, as noted by the evaluators, were presented previously in the results section. A review of the deficiencies revealed specific types of deficiencies characteristic of Mandate Consultant and human experts, respectively. The deficiences for Mandate Consultant were grouped primarily into two categories: (a) conclusions noted as deficient because the evaluator interpreted regulations differently than the knowledge base of Mandate Consultant, and (b) conclusions noted as deficient because the special educators who used Mandate Consultant interpreted cumulative file information differently than the evaluator. Neither of these deficiencies necessarily indicated an inaccuracy in the knowledge base of the system. Evaluators sometimes interpreted the regulations less explicitly. For example, Mandate Consultant noted the failure of an IEP to identify a timeline for an annual review. However, the evaluator commented that even though an anticipated annual review date was missing there was no reason to assume that the local education agency would not meet the annual review requirement. Also, the difference between the special educator's and the evaluator's interpretation of cumulative file data did not suggest error in the knowledge base of Mandate Consultant. Rather, it raised a different, but relevant, issue regarding the accuracy of the data provided by the user during a consultation with Mandate Consultant.

The deficiences noted for the human expert conclusions were grouped into three general categories: (a) an insufficient number of conclusions were identified for one to make data-based judgments about the case; (b) conclusions did not address specific procedural requirements of the regulations, for example, failure to note lack of appropriate IEP team members or a failure to state goals for related services on the IEP; and (c) conclusions were subjective, that is, they were based on opinion and presumption rather than on the procedural requirements of the regulations. The last finding, regarding opinion-based conclusions judged as deficient by evaluators, agreed with a conclusion by Colbourn (1982). She noted that the diagnostic reports written by humans often contained subjective impressions not present in the reports generated by the expert system.

The deficiencies noted in the conclusions of human experts supported the need for a knowledgeable, objective consultant such as Mandate Consultant for school administrators. Other evidence from this study supported Mandate Consultant as capable of generating valid information comparable to, and exceeding in many cases, the information generated by human experts.

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The IEP process serves as a forum during which parents and school officials should reach agreement on the content and provision of a handicapped child's education. When this process fails, other parties intervene to mediate the disagreement. If mediation is unsuccessful, the parties involved proceed to a hearing to resolve the issue based on the intent of the law. Unfortunately, hearings delay services to the student, cost much in terms of money and stress, and do not insure equitable, effective decisions. Clearly, school officials and parents need an objective, knowledgeable expert to review problems regarding IEP development so conflict may be resolved prior to formal due process hearings.

In recent years in the fields of medicine, geology, and engineering, specific domains of knowledge possessed by humans have been cloned in computer-based expert systems. However, few expert systems have been developed in the field of education. Therefore, the purpose of this dissertation was the development and initial validation of a microcomputer-based expert system designed to review the regulatory procedures implemented for IEP development.

The development of the expert system followed a Research and Development (R&D) model consisting of four phases. The first three phases emphasized formative evaluation. They included:

1. Product definition: During this phase the author identified objectives for the development of the expert system.
2. Product design: The design phase involved operationalizing outcomes of the definition phase into specific procedures for developing the expert system.
3. Product prototype: In this phase, a prototype of the expert system was developed and then systematically tested and revised until it produced reliable output. When the expert system yielded consistent results, that is, when more than 80 percent of the system's conclusions agreed with the conclusions of formative evaluators, a summative evaluation was conducted.

The fourth phase of the R&D model involved a summative evaluation designed as an initial validation of the conclusions produced by the expert system. The purpose of the evaluation was to assess whether the knowledge-based system accurately emulated the knowledge of experienced educators. The experimental design involved two formal evaluation components. In the first phase, six human experts, from a pool of experts identified by the staff of the state education agency, reviewed the data of ten field-representative cumulative student files and provided conclusions regarding failures to implement federal and state regulatory procedures for IEP development. In addition, conclusions about the IEP procedures were generated from the same ten cases by consultations with the expert system. The conclusions from the expert system and human experts were compared.

In the second evaluation component, three additional human expert evaluators reviewed the conclusions and judged their acceptability using a four-category rating scale: (a) 1-ideal, (b) 2-acceptable, (c) 3-less than acceptable, and (d) 4-unacceptable. These reviewers did not know the source of the conclusions, including not knowing that one of the sources was a computer program. The evaluators' ratings served as a basis for determining to what degree conclusions from the expert system and human experts were judged acceptable.

The summative evaluation yielded the following findings:

1. The conclusions produced by consultations using Mandate Consultant generally matched the conclusions of the "better" human experts and exceeded the conclusions of the majority of human experts. Consultations using Mandate Consultant produced the highest number of conclusions (56), the greatest number of interexpert agreements

(76), and the most frequently-noted conclusions for the cases (20). Kendall's coefficient of rank correlations (Tau) were computed on the rankings of the experts based on (a) the total number of conclusions and interexpert agreements, (b) the total number of interexpert agreements and most frequently-noted conclusions, and (c) the total number of conclusions and most frequently-noted conclusions. All three correlations were statistically significant at the .01, .03, and .03 levels, respectively.

2. Generally the evaluators judged the conclusions produced by consultations using Mandate Consultant as acceptable as those produced by the "better" human experts and more acceptable than the majority of human experts. Mandate Consultant, along with Human Expert 2, was rated as having the highest percentage (60%) of case reports judged acceptable by the majority of evaluators. Kendall's coefficient of concordance correlation (W), computed on the rankings of the expert case reports judged acceptable by the respective evaluators, was statistically significant at the .05 level.

Notably, the behavior of the human experts varied widely. Two of the human experts identified many conclusions after reviewing the ten cumulative student files, while four human experts identified few conclusions. In general, the case reports of these four human experts were judged less than acceptable or unacceptable by the evaluators. This information suggested two additional conclusions: (a) Educators may expect experienced special education administrators to demonstrate more expertise in reviewing the procedures of IEP development than they often do; and (b) if very few special educators possess the expertise required to accurately review the procedures of IEP development, the need for an unbiased, knowledgeable resource, such as Mandate Consultant, is even greater than originally anticipated by the author.

Conclusions

An analysis of the data reported in this study suggested the following conclusions regarding the output of the expert system:

1. The comparison of the conclusions produced by human experts and Mandate Consultant in terms of the number of conclusions, interexpert agreements, and most frequently-noted conclusions provides evidence that the knowledge base of a microcomputer-based expert system can emulate the knowledge of human special education

administrators regarding regulatory procedures for IEP development.

2. Independent evaluators' ratings of case reports produced by human experts and Mandate Consultant provide evidence that consultations using the expert system can produce conclusions that are as acceptable as those produced by human experts.
3. Analyzing the performance of the majority of the experienced special education administrators involved in Phase I indicates that many human experts may not demonstrate the level of skill anticipated regarding their review of regulatory procedures for IEP development.

Three other conclusions resulting from the experience of developing an expert system in special education included:

1. The application of artificial intelligence, specifically microcomputer-based expert systems, to special educators' administrative needs appears feasible.
2. The formative evaluation model suggested by Hofmeister (in press) provides a useful, systematic approach to expert system development in the field of education.
3. The "double-blinded" experimental design used for expert system evaluation in the field of medicine

is appropriate for evaluating educational expert systems.

Recommendations

The results of this study lead to the following recommendations:

1. Mandate Consultant needs to be expanded both in breadth and depth, that is, program code should be written to cover additional areas addressed by the regulations, such as referral and evaluation. Also, the current program could be expanded to address more qualitative issues, such as the instructional content of IEP goals and objectives.
2. The overall acceptability of Mandate Consultant in a field setting still requires evaluation. This study evaluated the validity of the output of the system, but additional evaluation should consider factors specific to user acceptance, such as accessibility, response time, and attitudes.
3. Mandate Consultant needs to be evaluated in terms of its potential as a training tool for the preservice and inservice of special educators. Several formative evaluators suggested that, based on their positive experiences with the system, Mandate Consultant should be adapted for training purposes.

4. The development and initial validation of Mandate Consultant should be viewed as only a beginning. Other expert system applications in the area of special education administration and instruction require further investigation.

REFERENCES

Ahmann, J. S. (1967). Aspects of curriculum evaluation: A synopsis. American Educational Research Association Monograph Series on Curriculum Evaluation, 1, 84-89.

Barr, A., & Feigenbaum, E. A. (Eds.). (1981). The handbook of artificial intelligence (1st ed.). Los Altos, CA: William Kaufman, Inc.

Belsches-Simmons, G., & Lines, P. (1984). Update: The legal rights of handicapped students. ECS Footnotes, 17(2), 4-7.

Borg, W. R., & Gall, M. D. (1983). Educational research: An introduction. New York: Longman, Inc.

Borich, G. D., & Jemelka, R. P. (1982). Programs and systems: An evaluation perspective. New York: Academic Press.

Brinkerhoff, R. O., Brethower, D. M., Hluchyj, Nowakowski, J. R. (1983). Program evaluation. Hingham, MA: Kluwer-Nijhoff Publishing.

Buchanan, B. G., Barstow, D., Bechtal, R., Bennett, J., Clancey, W., Kulikowski, C., Mitchell, T., & Waterman, D. A. (1983). Constructing an expert system. In F. Hayes-Roth, D. A. Waterman, & D. B. Lenat (Eds.), Building expert systems (pp. 127-168). Reading, MA: Addison-Wesley Publishing.

Budoff, M., Orenstein, A., & Abramson, J. (1981). Due process hearings: Appeals for appropriate public school programs. Exceptional Children, 48, 181-182.

Colbourn, M. (1982). Computer-guided diagnosis of learning disabilities: A prototype. Unpublished master's thesis, University of Saskatchewan, Saskatoon.

Cronbach, L. J. (1982). Designing evaluation of educational and social problems. San Franciso, CA; Josey-Bass.

Davis, R., Buchanan, B., & Shortliffe, E. (1975). Production rules as a representation for a knowledge-based consultant program. (Report STAN-CS-75-519 Memo AIM 226). Stanford, CA: Stanford University, Computer Science Department.

DeSalvo, D. A., Glamm, A. E., & Liebowitz, J. (1985). Structured design of an expert system prototype at The National Archives (Technical Paper). Arlington, VA: American Management Systems, Inc., & Washington, D. C: George Washington University.

DeSalvo, D. A., & Liebowitz, J. (1985). The application of an expert system for information retrieval at The National Archives. (Technical Paper). Arlington, VA: American Management Systems, Inc., & Washington, DC: George Washington University.

Duda, R. O. & Shortliffe, E. H. (1983). Expert systems research. Science, 220, 261-268.

Ekstrand, R. E. (1979). Preparing for the due process hearing: What to expect and what to do. Amicus, 4(2), 91-96.

Evans, S. J., Norwich, K. H., Cobbold, R. S. C., Diehl, D. L., Haberman, H., Harvey, B. J., O'Beirne, H., & Zingg, W. (1984). Computer diagnosis of skin disease: System design and preliminary results. International Journal of Biomedical Computing, 15, 271-284.

Ferguson, G. A. (1981). Statistical analysis in psychology and education (5th ed.). New York: McGraw-Hill.

Ferrara, J. M., & Hofmeister, A. M. (1984). CLASS.LD: An expert system for classifying learning disabilities [Computer program]. Logan, UT: Utah State University.

Fiedler, C. (1985). Determinants of due process conflict. Unpublished doctoral dissertation, University of Kansas, Lawrence.

Gaschnig, J. (1982). Application of the PROSPECTOR system to geological exploration problems. In J. E. Hayes, D. Michie, & Y. H. Pao (Eds.), Machine intelligence (pp. 301-323). New York: John Wiley & Sons, Inc.

Gaschnig, J., Klahr, P., Pople, H., Shortliffe, E., & Terry, A. (1983). Evaluation of expert systems: Issues and case studies. In F. Hayes-Roth, D. A. Waterman, & D. B. Lenat (Eds.), Building expert systems (pp. 241-282). Reading, MA: Addison-Wesley Publishing.

Goodall, A. (1985). The guide to expert systems. Oxford, England: Cotswold Press.

Harmon, P., & King, D. (1985). Artificial intelligence in business: Expert systems. New York: John Wiley & Sons, Inc.

Hemphill, J. K. (1969a). Educational research, educational development, and evaluation studies. In J. K. Hemphill (Ed.), Educational development: A new discipline for self-renewal (pp. 34-39). Eugene, OR: Center for the Advanced Study of Educational Administration.

Hemphill, J. K. (1969b). The relationships between research and evaluation studies. In J. K. Hemphill (Ed.), Educational development: A new discipline for self-renewal (pp. 149-171). Eugene, OR: Center for the Advanced Study of Educational Administration.

Henderson, R. A., & Hage, R. E. (1979). Economic implications of public education of the handicapped. Journal of Research and Development in Education, 12(4), 71-79.

Hickam, D. H., Shortliffe, E. H., Bischoff, M. B., Scott, A. C., & Jacobs, C. D. (in press). A study of the treatment advice of a computer-based cancer chemotherapy protocol advisor. Annals of Internal Medicine.

Hofmeister, A. M. (in press). Formative evaluation in the development and validation of expert systems in education. Computational Intelligence.

Hofmeister, A. M. & Ferrara, J. M. (in press). Expert systems and special education. Exceptional Children.

Hofmeister, A. M., & Lubke, M. (in press). Expert systems: Implications for the diagnosis and treatment of learning disabilities. Learning Disability Quarterly.

Hood, P. D. (1973). Evaluation stages for major laboratory products. In J. K. Hemphill (Ed.), Educational development: A new discipline for self-renewal (pp. 101-107). Eugene, OR: Center for the Advanced Study of Educational Administration.

Kammerlohr, B., Henderson, R. A., & Rock, S. (1983). Special education due process hearings in Illinois. Exceptional Children, 49, 417-422.

Kidd, A. (1984). Human factors problems in the design and use of expert systems. In A. Monk (Ed.), Fundamentals of human-computer interaction (pp. 237-247). Orlando, FL: Academic Press.

Lehr, D. & Haubrich, P. (1986). Legal precedents for students with severe handicaps. Exceptional Children, 52(4), 358-365.

Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., Lederberg, J. (1980). Applications of artificial intelligence for organic chemistry: The DENDRAL Project. New York: McGraw-Hill.

Markle, S. M. (1967). Empirical testing of programs. In P. C. Lange, (Ed.), Programmed instruction (pp. 104-140). Sixty-sixth yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Martuza, V.R. (1977). Applying norm-referenced and criterion referenced measurement in education. Boston, MA: Allyn & Bacon, Inc.

McDermott, J., (1982). R1: A rule-based configurer of computer systems. Artificial Intelligence, 19, 39-88.

Meadow, C. T., Hewett, T. T., & Aversa, E. G. (1982a). A computer intermediary for interactive data base searching: One design. Journal of the American Society for Information Science, Sep., 325-332.

Meadow, C. T., Hewett, T. T., & Aversa, E. G. (1982b). A computer intermediary for interactive data base searching. Journal of the American Society for Information Science, Nov., 357-364.

Meyers, W. R. (1981). The evaluation enterprise. San Francisco: Josey-Bass.

Miller, R. A., Pople, H. E., Jr., & Myers, J. D. (1982). Internist-I, an experimental computer-based diagnostic consultant for general internal medicine. The New England Journal of Medicine, 307, 468-476.

Mulsant, B., & Servan-Schreiber, D. (1984). Knowledge engineering: A daily activity on a hospital ward. Computers and Biomedical Research, 17, 71-91.

Nevo, D. (1983). The conceptualization of educational evaluation. Review of Educational Resesarch, 53(1), 117-128.

Othman, M., & Seireg, A. (1983). Expert system goes underwater. Computers in Mechanical Engineering, 2, 22-29.

Parry, J. D., & Ferrara, J. M. (1985). The potential of computer-based expert systems for special educators in rural settings. Small School Forum, 7(1).

Patton, M. Q. (1980). Qualitative evaluation methods. Beverly Hills, CA: SAGE Publications.

Politakis, P. & Weiss, S. M. (1984). Using empirical analysis to refine expert system knowledge bases. Artificial Intelligence, 22, 23-48.

Rychener, M. D. (1985). Expert systems in engineering design. Expert Systems, 2, 30-41.

Salend, S. J., & Zirkel, P. A. (1984). Special education hearings: Prevailing problems and practical proposals. Education and Training of the Mentally Retarded, 19(1), 29-34.

Sanders, J. R., & Cunningham, D. J. (1974). Formative evaluation: Selecting techniques and procedures. In G. D. Borich (Ed.), Evaluating educational programs and products (pp. 279-311). Englewood Cliffs, NJ: Educational Technology Publications.

Scriven, M. (1967). The methodology of evaluation. Perspectives of curriculum evaluation. AERA Monograph Series, Chicago: Rand McNally.

Scriven, M. (1973). The methodology of evaluation. In B. R. Worthen & J. R. Sanders (Eds.), Educational evaluation: Theory and practice. Belmont, CA: Wadsworth Publishng.

Scriven, M. (1981). Product evaluation. In N. L. Smith (Ed.), New techniques for evaluation (pp. 121-166). Beverly Hills, CA: SAGE Publications.

Shortliffe, E. H., (1981). Consultation systems for physicians: The role of artificial intelligence techniques. In B. L. Webber & N. J. Nilsson (Eds.), Readings in artificial intelligence (pp. 323-333). Palo Alto, CA: Tioga Publishing Company.

Siegel, A. (1956). Nonparametric statistics: For the behavioral sciences. New York: McGraw-Hill.

Sleeman, D. & Brown, J. S. (1982). Introduction: Intelligent tutoring systems. In D. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems (pp. 1-11). New York: Academic Press.

Smith, N. L. (1981). Developing evaluation methods. In N. L. Smith (Ed.), Metaphors for evaluation (pp. 17-49). Beverly Hills, CA: SAGE Publications.

Smith, T. E. C. (1981). Status of due process hearings. *Exceptional Children*, *48*, 232-236.

Spencer, R. H. (1985). *Computer usability testing and evaluation*. Englewood Cliffs, NJ: Prentice-Hall.

Stufflebeam, D. L. (1974). Alternative approaches to educational evaluation: A self-study guide for educators. In W. J. Popham (Ed.) *Evaluation in education*. Berkeley, CA: McCutchan Publishing.

Turnbull, A. P., Strickland, B., & Turnbull, H. R. (1981). Due process hearing officers: Characteristics, needs, and appointment criteria. *Exceptional Children*, *48*, 48-54.

Turnbull, H. R., Turnbull, A. P., & Strickland, B. (1979). Procedural due process: The two-edged sword that the untrained should not unsheath. *Boston University Journal of Education*, *161*(3), 40-59.

Waterman, D. A. (1986). *A guide to expert systems* (pp. 224-226). Reading, MA: Addison-Wesley Publishing.

Waterman, D. A., & Hayes-Roth, F. (1983). An investigation of tools for building expert systems. In F. Hayes-Roth, D. A. Waterman, & D. B. Lenat (Eds.), *Building expert systems* (pp. 169-218). Reading, MA: Addison-Wesley Publishing.

Weiss, S. M., Kulikowski, C. A., & Amarel, S. (1978). A model-based method for computer-aided medical decision making. *Artificial Intelligence*, *11*, 145-172.

Wills, K., Teather, D., & Innocent, P. (1982). An expert system for the medical diagnosis of brain tumors. International Journal of Man-Machine Studies, 16, 341-349.

Worthen, B. R. (1979). Eclecticism & evaluation models: Snapshots of an elephant's anatomy (Technical Paper). Logan, UT: Utah State University.

Wraith, S. M., Aikins, J. S., Buchanan, B. G., Clancey, W. J., Davis, R., Fagan, L. M., Hannigan, J. F., Scott, A. C., Shortliffe, E. H., VanMelle, W. J., Yu, V. L., Axline, S. G. & Cohen, S. N. (1976). Computerized consultation system for selection of antimicrobial therapy. American Journal of Hospital Pharmacy, 33, 1304-1308.

Yu, V. L., Buchanan, B. G., Shortliffe, E. H. Wraith, S. M., Davis, R., Scott, A. C., & Cohen, S. N. (1979). Evaluating the performance of a computer-based consultant. Computer Programs in Biomedicine, 9, 95-102.

Yu, V. L., Fagan, L. M., Wraith, S. M., Clancey, W. J., Scott, A. C., Hannigan, J., Blum, R. L., Buchanan, B. G., & Cohen, S. N. (1979). A blinded evaluation by infectious disease experts. Journal of the American Medical Association, 242, 1279-1282.

Zirkel, P. A. (1983). Building an appropriate education from "Board of Education v. Rowley": Razing the door and raising the floor. Maryland Law Review, 42, 466-495.

APPENDICES

Appendix A

Table 16

Summary of Studies Addressing the Evaluation of Computer-based Expert Systems

Table 16

Summary of Studies Addressing the Evaluation of Computer-based Expert Systems

EXPERT SYSTEM	SOURCE	FIELD	TYPE OF EVALUATION	ABSTRACT OF EXPERT SYSTEM EVALUATION
CASNET - designed to assist with the diagnosis and treatment of glaucomas.	Weiss, S.M., Kulikowski, C.A., & Amarel, S. (1978)	Medicine	Informal evaluation of the accuracy of the knowledge base.	Ophthalmologists attending an annual meeting of the American Academy of Ophthalmology and Otolaryngology that included an exhibit of CASNET were invited to present difficult cases to the computer. 77% of the ophthalmologists who entered difficult cases into CASNET described it as performing at an expert or very competent level.
MYCIN - provides antimicrobial therapy recommendations for treatment of bacterial infections.	Yu, V.L., Buchanan, B.G., Shortliffe, E.H., Wraith, S.M., Davis, R., Scott, A.C., & Cohen, S.N. (1979).	Medicine	Formal evaluation of the accuracy of the knowledge base.	Study involved 10 infectious disease specialists (5 from Stanford and 5 from other locations). Each expert evaluated medical data on each of a set of 15 patients and made recommendations for treatment. The human expert recommendations were compared to the expert system recommendations. Results of the evaluation show that the program's therapy recommendations meet Stanford University experts' standards of acceptable practice 91.9% of the time.
MYCIN - provides antimicrobial therapy recommendations for treatment of bacterial infections.	Yu, V.L., Fagan, L.M., Wraith, S.M., Clancey, W.J., Scott, A.C., Hannigan, J., Blum, R.L., Buchanan, B.G., & Cohen, S.N. (1979).	Medicine	Formal evaluation of the accuracy of the knowledge base. This study was a "blinded" evaluation that built on the results of the earlier study by Yu, Buchanan, et al. (1979).	Eight independent evaluators with special expertise in the management of meningitis compared Mycin's choice of antimicrobials with the choices of 9 human prescribers for 10 cases of meningitis. Mycin received an acceptability rating of 65% by the expert evaluators, the corresponding acceptability of human experts as rated by expert evaluators ranged from 42.5 to 62.5%.
DENDRAL - a group of programs designed to aid organic chemists with interpret data from samples of unknown compounds.	Lindsay, R.K., Buchanan, B.G., Feigenbaum, E.A., & Lederberg, J. (1980).	Chemistry	Informal and formal evaluations have been completed on several families of the Dendral system, e.g., Congen, Planner, Mollen, and Meta-Dendral.	A series of tests were completed following a progression from relatively simple compounds to more complex compounds. Rules were developed based on a small number of compounds in a class. These rules were modified and expanded as additional compounds were added. Then the system was run on other compounds in the same class to test the program's accuracy and generality. The positive results from numerous tests build a strong case for the range and power of Dendral to generate accurate explanations of experimental data.
PROSPECTOR - evaluates the mineral potential of a site or region with respect to specific classes of ore deposits.	Gaschnig, J. (1982).	Geology	Formal evaluations of the accuracy of the knowledge base as well as the acceptability of the system were conducted. Results reported here address the evaluation of the accuracy of the knowledge base.	For eight geographical regions, a relative comparison was made between the conclusions of a Prospector model and those of the expert geologist who designed it. Prospector's detailed conclusions matched the designer's conclusions to within about 7% on a 10-point certainty scale used as a basis for comparison.
IIDA (Individualized Instruction for Data Access System) - designed to aid scientific and technical end users of online bibliographic information retrieval systems in the performance of database searches.	Meadow, C.T., Hewett, T.T., & Aversa, E.S. (1982a).	Library Science	Formative and summative evaluations were completed. Formative evaluation feedback and information were used for refinement and further development of the system. The summative evaluation assessed the impact and effectiveness of the system based on the project objectives. A formal evaluation design is reported only for the summative portion.	For a baseline study as well as a training method study, 50 participants were selected with half of the participants being randomly assigned to each group (control and experimental). No statistically significant differences were found between the two groups in either study. The system was tested in an industrial setting and it was demonstrated that (1) end users of scientific and technical literature could learn to do their own bibliographic searches through computer-assisted instruction as well as they learned through a comparable period of conventional instruction, and (2) end users were as satisfied with the results of their own searchings as with the result of searches performed for them, for the types of searches tested.

BEST COPY AVAILABLE

Table 16 (Continued)

EXPERT SYSTEM	SOURCE	FIELD	TYPE OF EVALUATION	ABSTRACT OF EXPERT SYSTEM EVALUATION'
INTERNIST-I - provides computer assisted diagnosis in general internal medicine.	Miller, R.A., Pople, H.E., Jr., Myers, J.D. (1982).	Medicine	First systematic evaluation of the accuracy of the knowledge base.	Nineteen cases were identified in which all the major diagnoses were already included in the database of Internist-I. Using these cases, diagnostic decisions made by the clinicians at Mass. General Hospital, by the case discussants (experts presented with the same data as presented to Internist-I), and by Internist-I were classified as correct when they were confirmed by the pathologists or when a clinical syndrome was universally agreed to be present. Of the 43 possible correct diagnoses in the 19 trial cases, Internist-I, the clinicians at the Mass. General Hospital, and the case discussants made 17, 23, and 29 correct definitive diagnoses, respectively.
DIAGNOSTIC ADVISOR - designed for the medical diagnoses of brain tumours.	Wills, K., Teather, D. & Innocent, P. (1982).	Medicine	Formal evaluation of the accuracy of the knowledge base.	Study involved comparison of Diagnostic Advisor diagnoses and radiologists' diagnoses for three types of tumour in 191 cases. Results of the evaluation shows that the program achieved an accuracy of 69% of its first choice in correctly diagnosing the disease, and an accuracy of 88% of its first three choices in correct diagnosis. These accuracy levels were similar to the panel of radiologists.
R1 - configures VAX-11/780 computer systems based on a customer order.	McDermott, J. (1982).	Computer Science	Formal evaluation of the accuracy of the knowledge base.	A team of 6 experts examined R1's output for 50 orders, spending from one to two hours on each order. Using R1 to configure computer systems for 50 orders, only 12 pieces of errorful knowledge were uncovered. These errors were corrected and R1 was judged to be expert enough to be used in place of human experts.
OCEAN CURRENT METER - prototype designed to control the movement of a free-falling current meter.	Othman, M. & Seireg, A. (1983).	Oceanography and Engineering	Simulated testing of expert system prototype.	Simulated tests were carried out to investigate the performance of the current meter. The tests performed showed that good accuracy can be expected.
COMPUTER-GUIDED DIAGNOSIS OF LEARNING DISABILITIES - designed as a computer based diagnostic system to assist educators in the assessment of learning disabled children.	Colbourn, M. J. (1983).	Education	Formal evaluation of the accuracy of the knowledge base of this prototype.	The output of the expert system prototype was compared to the diagnoses of two human experts for 24 cases. Lack of agreement between the two human experts made it difficult to judge the expert system's performance. Nonetheless, the results demonstrated that the expert system's diagnostic reports tend to be both more accurate and more extensive than those of human diagnosticians.
BLUE BOX PROJECT - designed for psychopharmacologic treatment of depression.	Mulsant, B. & Servan-Schreiber, D. (1984).	Medicine	Informal assessment of knowledge base accuracy. Study also addressed acceptability of the system from the perspectives of patients and medical students.	Informal assessment using discharge summaries of former patients found a general consistency between the advice of the program and the discharge plan.
DIAG - designed for the medical diagnosis of dermatological disease.	Evans, S.J., Horwich, K.H., Cobold, R.S.C., Diehl, D.L., Heberman, H., Harvey, D.J., O'Beirne, M., & Ziegg, W. (1984).	Medicine	Formal preliminary evaluation of the accuracy of the knowledge base.	Study involved comparison of DIAG diagnoses and an experienced dermatologist's diagnoses for 50 cases. Results of the evaluation indicate that while the program nearly always includes the correct diagnosis in its differential diagnosis (94% of the cases), it frequently also includes diseases thought to be inappropriate (34% of the diseases listed).

Compiled by James D. Parry (1985)

BEST COPY AVAILABLE

Appendix B

Figure 5

An Analysis of the Studies Addressing Computer-based Expert Systems

	WHOM IS ES INTENDED?			WHAT IS BEING EVALUATED?					WHAT WERE THE OUTCOMES?			WHAT ASSESSMENTS AND COMPARISONS WERE MADE?					WHO WILL EVALUATE THE RESULTS?						WERE THREATS TO VALIDITY CONTROLLED FOR? (1=high 5=low)					HOW WERE TEST CASES SELECTED?			WHAT IS THE PURPOSE OF THE EVALUATION? Formative or Summative	
	Developers of System	Experts End Users	Researchers	Quality of Decisions Accuracy	Correctness of Reasoning	Quality of Human-Computer Interaction	System's Efficiency	Cost Effectiveness	H (40-80)	M (20-40)	L (0-20)	Human's Assessed	ES Assessed	Accuracy	Acceptability	Educational Potential	Human Expert	End User	Developer of System	Individuals	Groups	By Consensus	1	2	3	4	5	Random	Prearranged Representative	NR	F	S
CASNET		EE		QD					77%			HE	ES	AT				EU		I							NR		P			?
MYCIN (1)		EE		QD					86.7% (All 10 Experts)			(Comparisons) HE Stanford Group HE Non-Stanford Group	ES ES	AT			HE (5 Stanford) (5 Non-Stanford)										5 (Undocumented)			NR (15)		?
MYCIN (2)		EE		QD					65%			HE	ES	AT		EP	HE (8) (9)	Stage 1 Stage 2		I		C		2 (Blinded Evaluation)						NR (10)		?
DENDRAL	DOS				CR					NR		"Most Widely Used and Accepted ES"			AT		HE		DOS								NR	I	P (Generalizability Tested)		?	
PROSPECTOR		EE		QD		QHCI			70%			(Comparison) HE	ES	AT					DOS								NR		P (8) (Geographic Regions)			?
IDA	DOS			QD *(Summative Phrase)		QHCI	SE*		"No Statistically Significant Difference"			Control and Experiment Groups Compared ES						EU		I				2 (Random Assignment To Groups)				R (50)			F	S
INTERNIST-I		EE		QD						39%		HE	ES	AT			HE	EU		I							NR	R (19)				?
DIAGNOSTIC ADVISOR				QD					60% (First Choice)			(Comparison) HE	ES	AT			HE			I							NR			NR (191)		?
R1	EE (Production)			QD					76%			HE	ES	AT			HE (6)										NR			(NR) 50 Orders	?	
OCEAN CURRENT METER	DOS	EE (Prototype)		QD		QHCI				NR		HE	ES		AT		HE		DOS								NR		"Several" Simulated Tests		?	
COMPUTER-GUIDED DIAGNOSIS OF LD	DOS	EE (Prototype)		QD						NR			ES		AT		HE (2)		DOS								NR			NR (24)		?
BLUE BOX PROJECT		EE		QD		QHCI				NR			ES	AT	AT	EP	HE										NR			NR (Discharge Summaries)	?	?
DIAG		EE		QD					94% (Overall Diagnosis) 34% (Inappropriate Diagnosis)				ES	AT			HE										NR			NR 50	?	

Figure 5. An Analysis of the Studies Addressing Computer-based Expert Systems

BEST COPY AVAILABLE

Appendix C

Figure 6

Product Definition Appraisal Plan

Figure 6. Product Definition Appraisal Plan

Evaluation Questions	Sub-Questions	Why Important	Procedure	Respondents/ Sample	Instrument	Criteria
1. Is there an established need for an expert system that would assist educators in the manner proposed?	a. Who needs the system? i.e., Who are the potential users? b. Why do they need the system? c. What must the system be like to meet the needs of potential users?	Provides a target audience for field testing and dissemination. Establishes the purpose of the system: goal production. Establishes performance requirements specific to the system - user interface that must be addressed in the design.	Questionnaires completed by a panel of experts. Experts review the following documents: A. Review of literature B. Description of the Expert System C. Evaluation preview Experts answer sub-questions on the questionnaire.	Utah SEA Staff AIRD Staff USU Faculty Hearing Officer Local Administration	Product Definition Questionnaire. The system developer will compile a list of answers to the questions. This list will be compared to actual responses of the panel.	Panel responses will correspond with system developer's conclusions with at least 80% agreement.
2. Is the need well documented and supported by data?	a. Is the review of the literature comprehensive? b. Do the conclusions in the review of the literature appear logical and well founded? c. Are the sources cited in the review of the literature credible? d. Do the conclusions of the review of the literature correspond with the experience of experts in the field? e. Can alternative hypothesis be concluded and defended?	a-d. Are necessary to establish the validity of the needs/ problem as cited. e. Is necessary to establish reliability of the needs interpretation.				
3. Does the proposed solution appear feasible in light of known constraints and resources?	a. Does the proposed system respond to the needs identified? (e.g., Do the goals and objectives correspond to the needs?) b. Is there adequate information from which to develop the proposed solution? c. Are there adequate resources (time, talent, money) to support development of the system? d. Does it appear that the proposed system is practical?	a. Establishes an initial examination of internal consistency and assure the product will have worth. b-c. Are necessary to establish confidence that development is possible. d. Is necessary to establish that the product is workable in the field.				

Appendix D

Description of Mandate Consultant

Description of Mandate Consultant

Description

Mandate Consultant (MC) will be a microcomputer-based expert system designed to provide school officials or parents with a review of the procedures implemented for the development of an IEP. The review could be considered a second opinion, or expert opinion, that is based on federal and state regulations. A microcomputer-based expert system, such as Mandate Consultant, should have the potential to provide school officials and parents with a relatively accessible, cost-effective, and knowledgeable source for obtaining advice. A flowchart illustrating the proposed function of MC appears in Figure 7.

The Mandate Consultant program will contain many if-then rules based on the procedures of the IEP process as stated in the federal and state regulations. These rules will comprise most of what is known as the knowledge base of the expert system. An accurate representation of IEP-process knowledge within the rules will provide the basis for knowledgeable output.

Programming Base of MC

The proposed version of Mandate Consultant will be developed using the authoring tool, M.1. M.1 is an expert system authoring tool available from Teknowledge, Inc., Palo Alto, CA.

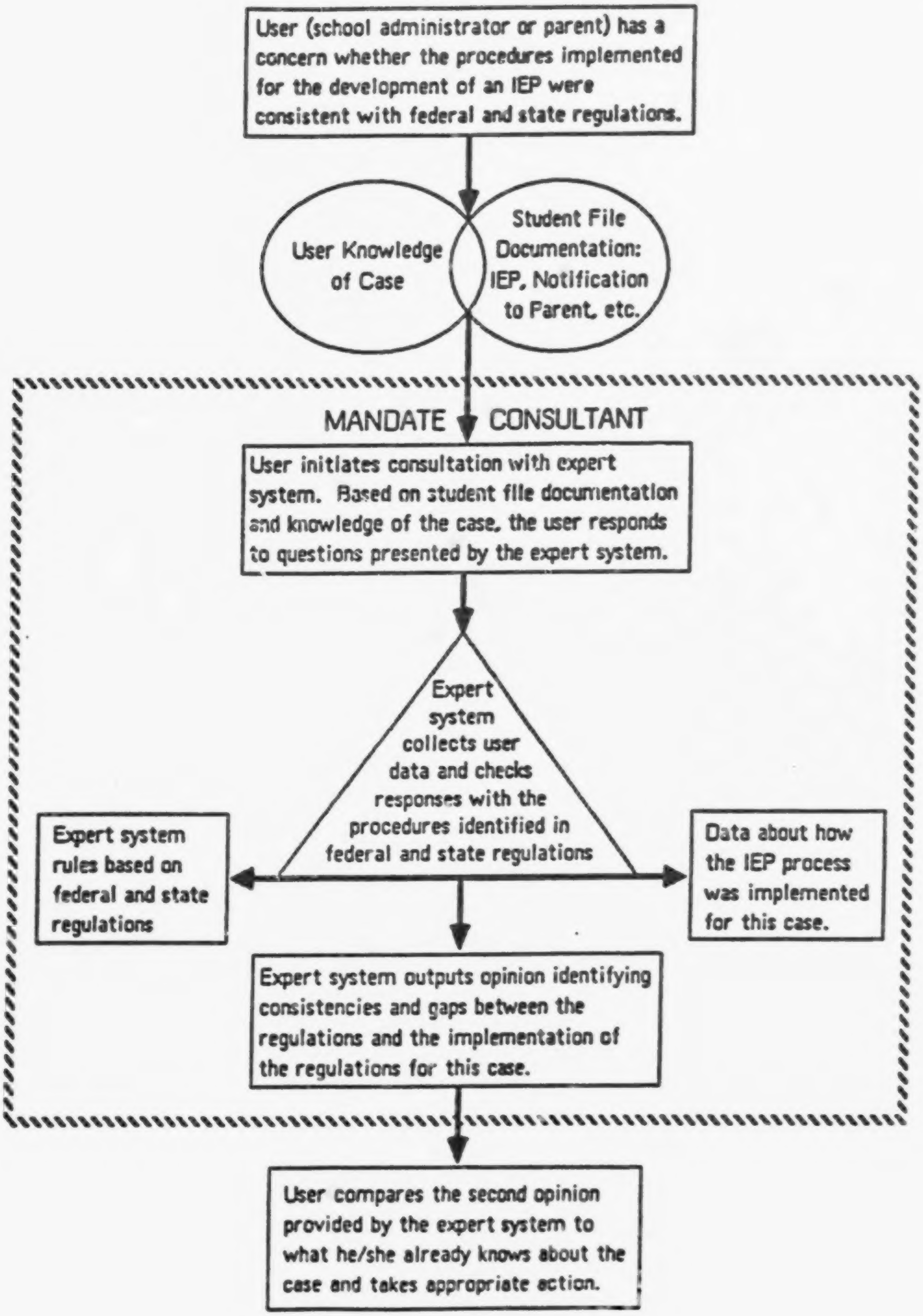

User (school administrator or parent) has a concern whether the procedures implemented for the development of an IEP were consistent with federal and state regulations.
User Knowledge of Case
Student File Documentation: IEP, Notification to Parent, etc.
MANDATE CONSULTANT
User initiates consultation with expert system. Based on student file documentation and knowledge of the case, the user responds to questions presented by the expert system.
Expert system collects user data and checks responses with the procedures identified in federal and state regulations
Expert system rules based on federal and state regulations
Data about how the IEP process was implemented for this case.
Expert system outputs opinion identifying consistencies and gaps between the regulations and the implementation of the regulations for this case.
User compares the second opinion provided by the expert system to what he/she already knows about the case and takes appropriate action.

Figure 7. Flowchart: How Mandate Consultant Operates

Equipment Requirements for MC

Mandate Consultant will run on an IBM PC or IBM-compatable computer. The proposed version of MC is designed to run on a 256K machine.

Appendix E

Evaluation Preview

for Mandate Consultant

Evaluation Preview
for
Mandate Consultant

The development of the expert system includes formative and summative evaluation components. The formative evaluation component will address: (1) product definition (clarification of need and proposed solution), (2) product design, and (3) product prototype. The summative evaluation will validate the accuracy of the output of the expert system. Brief descriptions of the evaluation components follow:

Product Definition

Purpose: Verify the identified need and proposed solution.

Primary Questions:

1. Is there an established need for an expert system that would assist educators in the manner proposed?
2. Is the need well documented and supported by data?
3. Does the proposed solution appear feasible in light of known constraints and resources?

Method: Group of experts read: (a) The Review of Literature, (b) The description of the expert system, (c) The evaluation preview, then respond to questionnaire.

Experts' responses will correspond with the system developer's conclusions with at least 80 percent agreement.

Product Design

Purpose: Verify that the design of the expert system is: (a) clear, (b) based on the needs data, and (c) compatible with constraints in the user's environment.

Primary Question:

1. Is the design of the expert system clear?
2. Is the design of the expert system internally consistent with the needs data?
3. Is the design of the expert system compatible with known constraints in the user's environment?

Method: Group of experts read: (a) The Review of the Literature, (b) Description of the Expert System, (c) Example of Expert System Consultation Output, (d) Federal and State Regulations, and (e) Translation of Expert System Knowledge Base. Selected experts complete a clarity evaluation instrument, internal consistency worksheet, and regulation checklist. In addition, two experts are interviewed about the output of the expert system. Experts' responses will correspond with the system developer's conclusions with at least 80 percent agreement.

Product Prototype

Purpose: Verify that the prototype of the expert system: (a) produces accurate output for a set of broad-based test cases and (b) can be successfully and independently operated by users.

Primary Questions:

A. Accuracy:

1. Are the questions asked by the system clear?
2. Are the responses, findings, and explanations from the system clear?
3. Does the system produce the intended conclusions?

B. Operation:

1. Can the intended user successfully operate the system using the materials and documentation provided?
2. Is the system's response time adequate?
3. Is it likely the intended user will be able to access the system?

Method: Selected experts will: (a) Review the documentation (e.g., IEP) from the test case files to identify inconsistencies between the procedures implemented for the test cases and the procedures outlined in the federal/state regulations, (b) Read the support materials (instructions/ documentation) about operating the expert system, (c) Operate the expert system using the test cases, (d) Compare their own expert conclusior about the test cases with the conclusions produced by the expert system, and (e) Complete questionnaires regarding the operation of the prototype. The experts' conclusions will correspond with expert system's conclusions on the test cases with at least 80 percent agreement. The experts' responses on the questionnaire will correspond with the system developer's conclusions with at least 80 percent agreement.

In addition, data will be collected to determine interuser reliability. The interuser reliability measure will exceed 80 percent agreement. Records of the consultations with the expert system using the test cases will be reviewed as a reliability check of the system's questions, responses, findings, and explanations. The data from the records of the consultations will correspond with at least 80 percent agreement.

Prototype Validation

Purpose: To determine if the output of a microcomputer-based expert system is accurate as compared to the knowledge of due process hearing officers.

Questions:

1. To what degree do expert system-generated decisions and human expert decisions match in terms of implementing or not implementing regulatory procedures for special education?
2. To what degree are expert system-generated decisions or human expert decisions rated as acceptable (based on a rating scale of 1-ideal, 2-acceptable, 3-less than acceptable, and 4-

unacceptable) by a panel of human expert evaluators?

Method: The design for this procedure involves two phases of a formal evaluation. In the first phase, the expert system will be used along with each of six human experts to review the data of ten case summaries and provide decisions regarding the implementation of regulatory procedures leading to an appropriate educational program. In the second phase, three additional human experts who do not know the source of the decisions (including not knowing that one of the sources was a computer program) each will review the decisions of the human experts and expert system and will judge the acceptability of the decisions using a rating scale. The evaluator's rating will serve as a basis for comparing the human expert and expert system decisions.

Appendix F

Questionnaire

for Product Definition Phase

Questionnaire
for Product Definition Phase

Upon reading the (a) Review of Literature, (b) Description of the Expert System, and (c) Evaluation Preview, please answer the subquestions that follow each of the three major questions.

1. Is there an established need for an expert system that would assist educators in the manner proposed?
    a. Who needs the system (i.e., who are the potential users?)?

    b. Why do they need the system?

    c. What must the system be like to meet the needs of potential users?

2. Is the need well documented and supported by data?
    a. Is the review of the literature comprehensive?

    b. Do the conclusions in the review of the literature appear logical and well-founded?

    c. Are the sources cited in the review of the literature credible?

d. Do the conclusions of the review of the literature correspond with your experience as an expert in the field?

e. Can alternative hypothesis be concluded and defended?

3. Does the proposed solution appear feasible in light of known resources and constraints?

a. Does the proposed system respond to the needs identified (e.g., Do the goals and objectives correspond to the needs?)?

b. Is there adequate information from which to develop the proposed solution?

c. Are there adequate resources (time, talent, money) to support development of the system?

d. Does it appear that the proposed system is practical?

Appendix G

Validity and Reliability of Product Definition Questionnaire

Validity and Reliability of Product Definition Questionnaire

Raw data from content specialists:

		Rater A	Rater B
Item #	1	4	4
	2	4	4
	3	4	4
	4	3	4
	5	3	4
	6	3	4
	7	4	4
	8	3	4
	9	4	4
	10	3	4
	11	4	3
	12	4	4

Coding of data using Martuza (1977) technique:

Rater A

		1 or 2	3 or 4	Totals
Rater B	1 or 2	0	0	0
	1 or 4	0	12	12
	Totals	0	12	12

4 = very relevant
3 = quite relevant
2 = somewhat relevant
1 = not relevant

Data analysis and results based on Martuza (1977) technique:

Interrater Agreement: Proportion of items given rating of (1 or 2) or (3 or 4) by both raters

$$\frac{\text{(Joint 1 or 2) + (Joint 3 or 4)}}{\text{Totals items}} = \frac{0 + 12}{12} = 1.0$$

Content Validity Index: Proportion of items given a rating of 3 or 4 by both raters

$$\frac{\text{Joint 3 or 4}}{\text{Total items}} = \frac{12}{12} = 1.0$$

Appendix H

Figure 8

Product Design Appraisal Plan

Appendix I

Clarity Evaluation Instrument for Product Design Phase

Clarity Evaluation

For Product Design Phase

I. Clarity of Objectives

Rate each of the objectives of the expert system twice. In the first rating, circle the number which best describes how clear the objective is. In the second rating, circle the letter which best summarizes the intention of the objective.

A. To provide school personnel and/or parents with an expert review of the process used to develop the IEP of a selected student.

	Confusing				Clear
1. Clarity of the objective	1	2	3	4	5

2. Intent of the objective
    - a. Improve parent-professional interaction
    - b. Provide an expert second opinion
    - c. Teach elements of federal/state law
    - d. Revise an IEP

B. To provide an economic alternative for seeking and obtaining expert advice on IEP development.

	Confusing				Clear
1. Clarity of objective	1	2	3	4	5

2. Intent of objective
    - a. Train IEP developers
    - b. Provide an expert second opinion

c. Revise local procedures

d. Reduce costs for consultation

II. Design

A. Examine the flow chart which describes the operation of the expert system. To what degree are the following elements clear in your mind?

		Confusing				Clear
1.	Who is expected to use system?	1	2	3	4	5
2.	What will the system produce?	1	2	3	4	5
3.	What input is necessary to run the system?	1	2	3	4	5
4.	What processes are performed on the input to produce the intended output?	1	2	3	4	5

B. Circle the answer which best describes or answers the following statement:

1. The expert system is primarily intended for use by...

a. Teachers

b. Hearing Officers

c. Parents

d. School Administrators

2. The expert system produces information that . . .

a. states how the IEP process could be improved.

b. provides alternative programming for use in the IEP.

c. states how the development of an IEP relates to the procedures identified in federal and state law.

d. provides legal advice which the school administrator can use in a subsequent hearing.

3. The data necessary to run the expert system are the:

a. user's responses to questions presented by the expert system.

b. IEP's.

c. teacher's observations of the student.

d. descriptions of problems with the case.

4. The opinion produced by the expert system results from:

a. a check of the user's responses.

b. a check of expert syustem rules based on state and federal regulations.

c. a check of the user's responses with the expert system rules based on state and federal regulations.

d. a check of the IEP with state and federal regulations.

Appendix J

Validity and Reliability of Clarity Evaluation Instrument

Validity and Reliability of Clarity Evaluation for Product Design Phase

Raw data from content specialists:

Ratings by:

Item #	Rater A	Rater B
1	4	4
2	4	4
3	4	4
4	4	4
5	4	4
6	4	4
7	4	4
8	3	4
9	4	4
10	4	4
11	4	4
12	3	4

Rating Scale

1 = not relevant
2 = somewhat relevant
3 = quite relevant
4 = very relevant

Coding of data using Martuza (1977) technique:

		Rater A		
		1 or 2	3 or 4	Totals
Rater B	1 or 2	0	0	0
	3 or 4	0	12	12
	Totals	0	12	12

Data Analysis and results based on Martuza (1977) technique:

Interrater Agreement: Proportion of items given rating of (1 or 2) or (3 or 4) by both raters

$$\frac{(\text{Joint 1 or 2}) + (\text{Joint 3 or 4})}{\text{Total items}} = \frac{0 + 12}{12} = 1.0$$

Content Validity Index: Proportion of items given a rating of 3 or 4 by both raters

$$\frac{\text{Joint 3 or 4}}{\text{Total items}} = \frac{12}{12} = 1.0$$

Appendix K

Sample of Mandate Consultant Program Code

Sample of Mandate Consultant Program Code

```
question(xx 'first time evaluation') = ['
Is this the first time the student was evaluated as a
potential special education student?   (yes or no)',nl,nl].
legalvals(xx 'first time evaluation') = [yes,no].


question(xx 'evaluation team member attend mtg') = ['
Did an evaluation team member or other knowledgeable person of the
student's evaluation attend the IEP team meeting?   (yes or no)',nl,nl].
legalvals(xx 'evaluation team member attend mtg') = [yes,no].


rule-105:
if xx 'first time evaluation' = yes
        and xx 'evaluation team member attend mtg' = yes
        then xx 'eval mtg' check = acceptable.


rule-110:
if xx 'first time evaluation' = yes
        and xx 'evaluation team member attend mtg' = no
        and display(['
Federal Regulation 300.344 and Utah State Regulation III.B. require
that if this is the first time the student has been evaluated as a
potential special education student, an evaluation team member or a','
person knowledgeable of the evaluation must participate in the IEP
team meeting.  Without this person, the IEP team is inconsistent with
the regulations.',nl,nl])
        then xx 'eval mtg' check = 'no eval team member' cf 90.


rule-115:
if xx 'first time evaluation' = no
        then xx 'eval mtg' check = acceptable.



/************************************************************/
question(xx 'private school placement') = ['
Is this student being considered for placement at a private school?
                                                   (yes or no)',nl,nl].
legalvals(xx 'private school placement') = [yes,no].


question(xx 'private school rep attend mtg') =['
Did a representative of the private school attend the IEP team
meeting?
                                                   (yes or no)',nl,nl].
legalvals(xx 'private school rep attend mtg') = [yes,no].


rule-120:
```

```
if xx 'private school placement' =  yes
        and xx 'private school rep attend mtg' = yes
        then xx 'private mtg' check = acceptable.


rule-125:
if xx 'private school placement' = yes
        and xx 'private school rep attend mtg' = no
        and display(['
Federal Regulation 300.347 and Utah State Rule III.B. require that
if a student is recommended for special education services at a
private school that a representative from the private school','
participate in the IEP meeting.  Without this representative, the
IEP team is inconsistent with the regulations.',nl,nl])
        then xx 'private mtg' check = 'no private school rep' cf 90.


rule-130:
if xx 'private school placement' = no
        then xx 'private mtg' check = acceptable.


/*****************************************************************/
question(xx 'student attend') = ['
Did the student attend the IEP team meeting?   (yes or no)',nl,nl].
legalvals(xx 'student attend') = [yes,no].


question(xx 'student appropriate') = ['
Would it have been appropriate for the student to attend the IEP
team meeting?   (yes, no, or unknown)',nl,nl].
legalvals(xx 'student appropriate') = [yes,no,unknown].


rule-135:
if xx 'student attend' = yes
        then xx student check = acceptable.


rule-140:
if xx 'student attend' = no
        and xx 'student appropriate' = yes
        and display(['
Federal Regulation 300.344 and Utah State Rule III.B. require
that the student should participate in the IEP team meeting when
appropriate.  For example, a secondary age student might be','
included in team discussions regarding his participation in a
vocational training program.  If the student has not participated
in the IEP team meeting, but his participation would have been','
appropriate, the team does not meet the requirements of the
regulations.',nl,nl])
        then xx student check = 'no student'.


rule-145:
```

```
if xx 'student attend' = no
        and xx 'student appropriate' = no
        or xx 'student appropriate' is unknown
        then xx student check = acceptable.


/*****************************************************************/
question(xx 'attend parochial') = ['
Is the student currently attending a parochial school or other
private school?   (yes or no)',nl,nl].
legalvals(xx 'attend parochial') = [yes,no].


question(xx 'parochial school rep attend mtg') = ['
Did a representative of the parochial school or other private
school attend the IEP team meeting?',nl,nl].
legalvals(xx 'parochial school rep attend mtg') = [yes, no].


question(xx 'alternative parochial rep involvement') = ['
If a representative could not attend, did the public agency
use other methods to insure participation of the private school
staff, including individual or conference telephone calls? ','
                                          (yes, no, or unknown)',nl,nl].
legalvals(xx 'alternative parochial rep involvement')
= [yes,no,unknown].


rule-150:
if xx 'attend parochial' = no
        then xx 'parochial mtg' check = acceptable.


rule-155:
if xx 'attend parochial' = yes
        and xx 'parochial school rep attend mtg' = yes
        then xx 'parochial mtg' check = acceptable.


rule-160:
if xx 'attend parochial' = yes
        and xx 'parochial school rep attend mtg' = no
        and xx 'alternative parochial rep involvement' = yes
        then xx 'parochial mtg' check = acceptable.


rule-165:
if xx 'attend parochial' = yes
        and xx 'parochial school rep attend mtg' = no
        and xx 'alternative parochial rep involvement' = no
        or xx 'alternative parochial rep involvement' is unknown
        and display(['
Federal Regulation 300.348 requires that if a handicapped child
is enrolled in a parochial or other private school and receives
special education or related services from a public agency,','
a representative of the parochial or private school should
participate in the IEP team meeting.  If the representative
cannot attend, the public agency is required to use other methods','
to insure participation by the private school, including individual
or conference telephone calls.  Without the involvement of a
representative, the IEP team does not meet the intent of the
regulations.',nl,nl])
        then xx 'parochial mtg' check = 'no parochial rep involvement'
```

Appendix L

Example of Expert System Output

Example of Expert System Output

```
What is the student's current age in years?

>> 15.

Was an IEP team meeting held for this student?   (yes or no)

>> y.

Did the student's parent(s) attend the IEP team meeting?
(yes or no)

>> y.

Did the public agency take whatever action was necessary to
insure the parent(s) understood the proceedings at the IEP
team meeting (for example, an interpreter for parents who
are deaf or whose native language is other than English)?
(yes, no, or unknown)

>> unknown.

Both the Federal Regulation 300.345(e) and the Utah State
Rule III.C.4. require that the public agency take the
necessary action to insure that the parent(s) understand the
proceedings at the IEP team meeting.


Did a representative of the public agency other than the
student's teacher attend the IEP team meeting?  (The
representative should be someone qualified to provide or
supervise the provision of special education services.)
(yes or no)

>> n.

Federal Regulation 300.344 and Utah State Rule III.B.
require that a public agency official other than the
student's teacher must participate in the IEP team meeting.
In addition, Utah state rules (III.B.1.) require that the
public agency representative must be authorized to commit
public agency resourses to provide special education
services for the student.  Without this public agency
representative, the IEP team does not meet the regulations.
```

```
Did the student's teacher attend the IEP team meeting?
(yes or no)

>> y.

Is this the first time the student was evaluated as a
potential special education student?   (yes or no)

>> y.

Did an evaluation team member or other knowledgeable person
of the student's evaluation attend the IEP team meeting?
(yes or no)

>> y.

Is this student being considered for placement at a private
school?  (yes or no)

>> n.

Did the student attend the IEP team meeting?   (yes or no)

>> n.

Would it have been appropriate for the student to attend the
IEP team meeting?   (yes, no, or unknown)

> unknown.

Is the student currently attending a parochial school or
other private school?   (yes or no)

>> n.

Is this the first time the student will be placed in
special education?   (yes or no)

>> y.
```

```
Did  the parent give approval by signature for placement  of
the student  in special education (that is,  signature on
the  parent permission for placement form)?   (yes or no)


>> y.

Did the IEP team develop a written IEP for this student?
(yes or no)


>> y.

Was the IEP developed within 30 days of when the student was
determined as eligible for special education services?
(Determination of eligibility is a multidisciplinary team
decision.  This team should include the parent whenever
possible.) (yes, no, or unknown)


>> y.

Does the IEP state the student's present levels of
educational performance?   (yes or no)


>> y.

How many of the present levels of performance statements are
based on the multidisciplinary observation and/or evaluation
data?  (all, some-statements, none, or uncertain)


>> a.

Does the IEP identify annual goals including short term
objectives for the student?   (yes or no)


>> y.

How many of the annual goals and short term objectives are
based on the unique needs of the student, that is, how many
of the goals and objectives are consistent with the findings
of the multidisciplinary evaluation?   (all, some, none, or
uncertain)


>> a.
```

How many of the short term objectives are measurable? (all, some, none, or uncertain)

>> s.

Do the annual goals or short term objectives describe the anticipated behaviors for the student to achieve? For example, the student will `add two digit numbers` or `remain in his seat. (yes, some-of-them, no, or uncertain)

>> s.

Does the IEP identify the specific special education services to be provided to the student? For example, resource assistance for math. (yes or no)

>> y.

Based on the multidisciplinary evaluation data, does the physical education program for this student require adaptation? (yes or no)

>> n.

Does this student require related services to support his/her special education services? For example, physical therapy or occupationaltherapy. (yes or no)

>> n.

How many of the service providers for the specific services stated on the IEP are identified? (all, some, or none)

>> a.

Does the IEP identify the extent to which the student will participate in the regular education program? (yes or no)

>> y.

Does the IEP identify the projected dates for the initiation of services? (yes, some-dates, no, or uncertain)

>> y.

Does the IEP identify the projected dates for the duration or termination of services? (yes, some-dates, no, or uncertain)

>> n.

Federal Regulation 300.346 and Utah State Rule III.D. require that the IEP state the projected dates for initiation of services and the anticipated duration of the services. Without this component, the IEP is incomplete.

Does the IEP identify criteria for evaluating student progress on program goals or objectives? For example, the student will add two digit numbers `correctly at 80% accuracy` or remain in his seat `for 20 consecutive minutes of the reading class period`. (yes, some-criteria, no, or uncertain)

>> y.

Does the IEP identify procedures for evaluating student progress on program goals or objectives? For example, testing the student with the Spencer Math Test may be the procedure for evaluating the student`s progress on math objectives, or teacher observation and record may be the procedure for evaluating if a student has remained seated for a specified time period. (yes, some-procedures, no, or uncertain)

>> y.

Does the IEP identify a schedule or timeline for at least an annual review of the student`s progress on his/her program? (yes, no, or uncertain)

>> n.

Federal Regulation 300.346 and Utah State Rule III.D. require that the IEP state appropriate schedules for determining, on at least an annual basis, whether the short term instructional objectives are being met. Without this component, the IEP is incomplete.

Was the IEP signed by all the IEP team members required by the regulations? (yes or no)

>> n.

Utah State Rule III.D. requires that each participant in the IEP team meeting sign and date the IEP. The participant's role or title also must be indicated. Without the signatures, the IEP is incomplete.

Was the IEP signed by the parent prior to placement and initiation of services for the student? (yes, no, or unknown)

>> y.

IEP Team Participation: The IEP team for this student failed to include all the participants required by the regulations. However, it is unknown if the public agency took the action necessary to insure that the parent understood the proceedings at team meeting.

Parent Approval of Placement: Acquisition of the parent's approval by signature for the proposed placement of the student in special education should meet the requirement of the regulations.

Development of an IEP: The development of an IEP by the IEP team within the thirty day limitation should meet the intent of the regulations.

Components of the IEP: The IEP has weak or missing components that are required by the regulations.

Team Approval of the IEP: The IEP was not signed by all of the appropriate IEP team members as required by the state rule.

Parent Approval of the IEP: Acquisition of parental consent by signature on the IEP prior to placement and initiation of services meets the requirements of the regulations.

Appendix M

Internal Consistency Worksheet

Internal Consistency Worksheet for Product Design Phase

Begin by completing column A. Enter in column A all the needs identified in the Review of the Literature provided (specifically, note "The Need" and "The Problem" sections of the Review). Use short phrases which capture the intent of each need you see identified. Next, complete column B based on the information in the Description of Mandate Consultant. Enter in column B short phrases which describe the intent of the objectives for the expert system. Finally, draw a line between the need phrases and the objective phrases which address those needs. It is not necessary that all needs or objectives be connected with lines. Only connect phrases with lines if you believe the objective is actually justified by the needs information.

Column A : Needs	Column B : Objectives

Now examine the product description. The objectives of the expert system are listed below. Under each objective list characteristics or products of the system that you think address the intention of the objectives. You may list as many items as you wish and you may decide not to list any items under an objective.

A. To provide school personnel and/or parents with an expert review of the process used to develop the IEP of a selected student.

____

____

____

____

B. To provide an economical alternative for seeking and obtaining expert advice on IEP development.

____

____

____

____

Appendix N

Regulations Checklist for Product Design Phase

Regulations Checklist
for Product Design Phase

The following is a checklist of federal and state regulations relevant to the IEP process. Review the knowledge base of the expert system (see the Translation of Expert System Code) and examine how each rule is used. Enter the code which best describes your opinion of the knowledge base.

N = The knowledge base never addresses this regulation in the rules.

C = This regulation is cited but not used in the rules.

U = This regulation is used but not as fully intended.

A = This regulation and the rules of the knowledge base agree totally.

____ 1. School districts in Utah are required to provide special education to eligible students from ages 5 through 21.

____ 2. IEP team meetings must be held for special education students.

____ 3. A representative of the school district other than the student's teacher must be involved in the IEP team meeting.

____ 4. The student's teacher must be involved in the IEP team meeting.

____ 5. The student's parent(s) should participate in the IEP team meeting.

____ 6. A surrogate parent may participate in the IEP team meeting if the child has no one fulfilling the role of parent for the IEP process.

____ 7. The surrogate parent may not be an employee of the school district that is responsible for the education of the student.

____ 8. A person knowledgeable of the student's evaluation must be involved in the IEP team meeting if this is the first time the student is being considered for special education placement.

____ 9. A representative of the private school should be involved in the IEP team meeting if the student is being considered for private school placement.

____ 10. When appropriate the student should be involved in the IEP team meeting.

____ 11. A representative of the parochial school or other private school should be involve in the IEP team meeting if the student is currently attending a private school.

____ 12. The public agency should take whatever action is necessary to insure that the parent(s) understood the proceedings at the IEP team meeting.

____ 13. The public agency must notify the parent(s) of an IEP team meeting for their child.

____ 14. The parent(s) must be notified of the IEP team meeting early enough to insure their opportunity to attend.

____ 15. The IEP team meeting must be scheduled at mutually agreed on time and place.

____ 16. The parent(s) must be notified of the purpose of the IEP team meeting.

____ 17. The parent(s) must be notified of the time of the IEP team meeting.

____ 18. The parent(s) must be notified of the location of the IEP team meeting.

____ 19. The parent(s) must be notified of who will be attending the IEP team meeting.

____ 20. The public agency must attempt to involve the parent(s) in the IEP team process through other means if the parent(s) is unable to attend the IEP team meeting.

____ 21. The public agency must document efforts to involve parent(s) in the IEP team meeting.

____ 22. The parent(s) must give their approval by signature for initial placement of their child or for significant changes in the child's program.

____ 23. The IEP team must develop a written IEP for each special education student.

____ 24. The IEP must be written within 30 days of determining the student is eligible for special education services.

____ 25. The IEP must state the student's present levels of educational performance.

____ 26. The present levels of performance must be based on multidisciplinary observation and/or evaluation data.

____ 27. The IEP must state annual goals including short term objectives.

____ 28. The annual goals must be based on the unique needs of the student.

____ 29. The short term objectives must be measurable.

____ 30. The annual goals and/or short term objectives must include criteria stating the anticipated behavior for the student to achieve.

____ 31. The IEP must state the specific special education services to be provided to the student.

____ 32. If the student requires an adaptive physical education program, the IEP must state the necessary modifications.

____ 33. If the student requires related services to support his/her special education program, the IEP must state these services.

____ 34. If the student requires related services to support his/her special education program, the IEP must state these services.

____ 35. The IEP must identify the service providers for the specific services stated on the IEP.

____ 36. The IEP must state the extent to which the student will participate in the regular education program.

____ 37. The IEP must identify the projected dates for initiation of services.

____ 38. The IEP must identify the projected dates for duration of services.

____ 39. The IEP must identify procedures for evaluating student progress on program objectives.

____ 40. The IEP must identify a timeline for at least an annual review of student's progress on his/her program.

____ 41. The IEP must be signed by all team members.

____ 42. The IEP must be signed by the parent prior to placement and initiation of services for the student.

Appendix O

Interview Form

for Product Design Phase

Interview Form
for Product Design Phase

After local administrators and/or hearing officers have reviewed examples of the output from the proposed expert system, collect their responses to the following questions.

1. Why might school administrators not use the output?

2. What circumstances surround these barriers?

3. How might the output be modified to increase the likelihood of its being used?

Appendix P

Validity and Reliability of Internal Consistency Worksheet for Product Design Phase

Validity and Reliability of Internal Consistency

Worksheet for Product Design Phase

Raw data from content specialists:

		Ratings by:	
		Rater A	Rater B
Item #	1	4	4
	2	4	4

Rating Scale

1 = not relevant
2 = somewhat relevant
3 = quite relevant
4 = very relevant

Coding of data using Martuza (1977) technique:

		Rater A		
		1 or 2	3 or 4	Totals
	1 or 2	0	0	0
Rater B	3 or 4	0	2	2
	Totals	0	2	2

Data analysis and results based on Martuza (1977) technique:

Interrater Agreement: Proportion of items given rating of (1 or 2) or (3 or 4) by both raters

$$\frac{(\text{Joint 1 or 2}) + (\text{Joint 3 or 4})}{\text{Total items}} = \frac{0 + 2}{2} = 1.0$$

Content Validity Index: Proportion of items given a rating of 3 or 4 by both raters

$$\frac{\text{Joint 3 or 4}}{\text{Total items}} = \frac{2}{2} = 1.0$$

Appendix Q

Validity and Reliability of Regulation Checklist for the Product Design Phase

Validity and Reliability of Regulation Checklist for the Product Design Phase

Raw data from content specialists:

Ratings for:

Item #	Rater A	Rater B
1	4	4
2	4	4
3	4	4
4	4	4
5	4	4
6	4	4
7	4	4
8	4	4
9	4	4
10	3	4
11	4	4
12	4	4
13	4	4
14	4	4
15	4	4
16	4	4
17	4	4
18	4	4
19	4	4
20	4	4
21	4	4
22	4	4
23	4	4
24	4	4
25	4	4
26	4	4
27	4	4
28	4	4
29	4	4
30	4	4
31	4	4
32	4	4
33	4	4
34	4	4
35	4	4
36	4	4
37	4	4
38	4	4
39	4	4

Rating Scale
1 = not relevant
2 = somewhat relevant
3 = quite relevant
4 = very relevant

		Rater A	Rater B
Item #	40	4	4
	41	4	4

Coding of data using Martuza (1977) Technique:

		Rater A		
		1 or 2	3 or 4	Totals
	1 or 2	0	0	0
Rater B	3 or 4	0	41	41
	Totals	0	41	41

Data analysis and results based on Martuza (1977) technique:

Interrater Agreement: Proportion of items given rating of (1 or 2) or (3 or 4) by both raters

$$\frac{\text{(Joint 1 or 2) + (Joint 3 or 4)}}{\text{Total items}} = \frac{0 + 41}{41} = 1.0$$

Content Validity Index: Proportion of items given a rating of 3 or 4 by both raters

$$\frac{\text{Joint 3 or 4}}{\text{Total items}} = \frac{41}{41} = 1.0$$

Appendix R

Validity and Reliability of Interview Form for Product Design Phase

Validity and Reliability of Interview Form for Product Design Phase

Raw data from content specialists:

Ratings by:

		Rater A	Rater B
Item #	1	4	4
	2	4	4
	3	4	4

Rating Scale
1 = not relevant
2 = somewhat relevant
3 = quite relevant
4 = very relevant

Coding of data using Martuza (1977) technique:

		Rater A		
		1 or 2	3 or 4	Totals
	1 or 2	0	0	0
Rater B	3 or 4	0	3	3
	Totals	0	3	3

Data analysis and results using Martuza (1977) technique:

Interrater Agreement: Proportion of items given rating of (1 or 2) or (3 or 4) by both raters

$$\frac{(\text{Joint 1 or 2}) + (\text{Joint 3 or 4})}{\text{Total items}} = \frac{0 + 3}{3} = 1.0$$

Content Validity Index: Proportion of items given a rating of 3 or 4 by both raters

$$\frac{\text{Joint 3 or 4}}{\text{Total items}} = \frac{3}{3} = 1.0$$

Appendix S

Figure 9

Prototype Appraisal Plan

Figure 9. Prototype Appraisal Plan

Evaluation Questions	Sub-Questions	Why Important	Procedure	Respondents/ Sample	Instrument	Criteria
Accuracy:						
1. Are the questions asked by the system clear?	a. Does the person operating the system understand what is being asked? / Are the questions confusing? b. Does each person operating the system interpret the questions in the same way?	This information is needed to establish if the questions presented by the system are understood by the users.	Selected experts will: (1) Review the documentation (e.g., IEPs) from ten test case files to identify inconsistencies between the procedures implemented for the test cases and the procedures outlined in federal/state regulations, (2) Read the support materials (instructions/documentation) about operating the expert system, (3) Operate the expert system using the test cases (4) Compare their own expert conclusions about the test cases with the conclusions produced by the expert system, and (5) Complete questionnaires regarding successful independent operation of the prototype.	Hearing Officer USU Faculty Regional Resource Center Staff AIRD Staff	Product Prototype Questionnaire: Parts A and B The system developer will compile a list of answers to the questions. This list will be compared to responses of the experts.	Experts' responses will correspond with system developer's conclusions with at least 80% agreement.
2. Are the responses, findings, and explanations from the system clear?	a. Does the person operating the system understand the responses, findings, and explanations provided by the system? / Are they confusing? b. Does each person operating the system interpret the responses and explanations in the same way?	This information is needed to establish if the responses, findings, and explanations by the system are understood by the user.				Experts' conclusions on test cases will correspond with conclusions produced by the expert system with at least 80% agreement. The reliability measures will exceed 80% agreement.
3. Does the system produce the intended conclusions?	a. How well do the conclusions produced by the system correspond with federal and state regulations? b. Is the system reliable, i.e., are the conclusions consistent when different users are given the same test case?	This information is needed to establish if the conclusions by the system consistently agree with the federal and state regulations.			Broad-based test cases. Record of differences between test case conclusions produced by experts and by the system.	
Operation:						
1. Can the intended user successfully operate the system using the materials and documentation provided?	a. Are the instructions/documentation logically organized? b. Are the instructions/documentation comprehensive? c. Are the instructions/documentation clear? d. Can users successfully operate the system using the instructions/documentation?	This information is needed to establish that the instructions/documentation are sufficient to provide for successful independent operation of the expert system.	Two or more experts will each run the same test cases and the conclusions produced by the system will be compared to evaluate interuser reliability.		Logs of each test case run by the experts using the expert system.	
2. Is the system's response time adequate?	a. Does the system respond quickly enough so the user won't be bored, distracted, or frustrated?	This information is needed to establish that the expert system can operate at a rate sufficient to meet the needs and expectations of the users.	The developer will record differences between the conclusions of the experts and the conclusions of the system. The developer will obtain and review the logs from each test case as a reliability check.			
3. Is it likely the intended user will be able to access the system?	a. Does the anticipated cost of the microcomputer hardware required to run the expert system make it accessible for the proposed user? b. Would the anticipated physical location of the microcomputer make it accessible for the users? c. Does the time required to run the expert system make its use feasible for the proposed users?	This information is needed to establish that the expert system will be sufficiently available to meet the needs of the intended users.				

Appendix T

Documentation for Operating Mandate Consultant: An Expert System for Application in Special Education

Documentation for Operating Mandate Consultant:

An Expert System for Application in Special Education

James D. Parry

Copyrighted 1985

Utah State University
Artificial Intelligence Research and Development Unit
Developmental Center for Handicapped Persons
Department of Special Education
UMC 6810, Logan, Utah 84321

Table of Contents

page

Description of Mandate Consultant (MC). 201

- Description. 201
- Programming Base of MC 201
- Equipment Requirements for MC 203
- Sample of a Consultation 203
- Example of "Explanation" Feature 203
- Example of Conclusions Produced by MC 204
- Current Version Emphasizes Development of the IEP 206

Instructions for Operating MC 206

- The Basics for Getting Started 206
- A. Typing Errors When Responding to Questions . 208
- B. Following User Responses with a Period . . . 208
- C. Auto-completion Feature for Entering User Answers 208
- D. Key Documents from Case Files Answering Questions by MC. 209
- E. A Record of the MC Consultation 209
- F. A Printed Copy of the MC Consultation 210
- G. Certainty Factors 211

Description of Mandate Consultant

Description

Mandate Consultant (MC) is a microcomputer-based expert system designed to provide school officials or parents with a review of the procedures implemented for the development of an IEP. The review could be considered a second opinion, or expert opinion, that is based on federal and state regulations. A microcomputer-based expert system, such as Mandate Consultant has the potential to provide school officials and parents with a relatively accessible, cost-effective, and knowledgeable source for obtaining advice. A flowchart illustrating the function of MC appears in Figure 10.

The Mandate Consultant program contains many if-then rules based on the procedures of the IEP process as stated in the federal and state regulations. These rules comprise most of what is known as the knowledge base of the expert system. An accurate representation of IEP-process knowledge within the rules provides the basis for knowledgeable output.

Programming Base of MC

The current version of Mandate Consultant was developed using the authoring tool, M.1. M.1 is an expert system authoring tool available from Teknowledge, Inc., Palo Alto, CA.

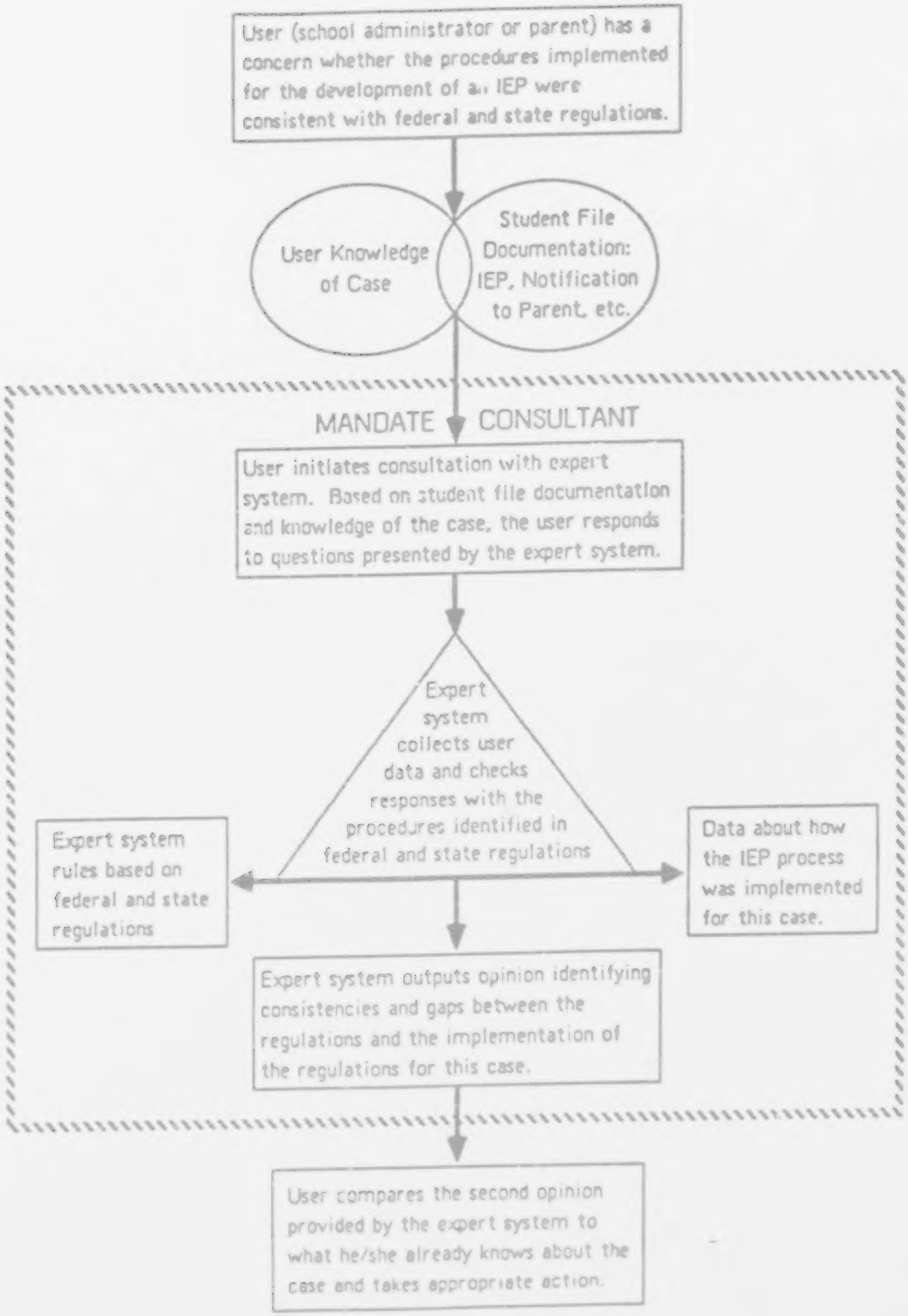

User (school administrator or parent) has a concern whether the procedures implemented for the development of an IEP were consistent with federal and state regulations.
User Knowledge of Case
Student File Documentation: IEP, Notification to Parent, etc.
MANDATE CONSULTANT
User initiates consultation with expert system. Based on student file documentation and knowledge of the case, the user responds to questions presented by the expert system.
Expert system collects user data and checks responses with the procedures identified in federal and state regulations
Expert system rules based on federal and state regulations
Data about how the IEP process was implemented for this case.
Expert system outputs opinion identifying consistencies and gaps between the regulations and the implementation of the regulations for this case.
User compares the second opinion provided by the expert system to what he/she already knows about the case and takes appropriate action.

Figure 10. Flowchart: How Mandate Consultant Operates

Equipment Requirements for MC

Mandate Consultant will run on an IBM PC or IBM-compatible computer. The current version of MC is designed to run on a 256K machine. With some modifications, MC will run on machines with more than 256K memory.

Sample of a Consultation

Following is a brief sample of the interaction between the user and MC. Depending on the user's input, MC may provide a response such as the one accompanying the second question of the sample.

Question presented by System - Did the student's teacher attend the IEP team meeting? (yes or no)

User's input - >> y.

Question presented by System - Did the student's parent(s) attend the IEP team meeting? (yes or no)

User's input - >> n.

Response from System - Federal Regulations 300.344 and 300.345, along with Utah State Rules III.B. and III.C., require that the parent participate in the IEP team meetings. Without the parent, the IEP team is inconsistent with the intent of the regulations.

Example of "Explanation" Feature

When the user is uncertain or unclear about why particular questions are presented by MC, an "explanation" feature of the system is helpful. If the user types "why."

in response to a question, the program presents an explanation informing the user why the information pursued by the question is relevant. An example showing the use of this feature follows:

Question presented by system - Does the IEP identify annual goals including short term objectives for the student? (yes or no)

User's input - >> why.

Explanation from system - Federal Regulation 300.345 and Utah State Rule III.D. require that the IEP include a statement of the student's annual goals including short term objectives. Without this component, the IEP is incomplete.

Example of Conclusions Produced by MC

When the system has collected the necessary data from the user to complete a review of the implementation of special education procedures for a student, the results of the review appear as the screen display and the consultation ends. The display includes eight summary statements with the level of certainty (see section entitled Certainty Factors) which may be attached to each element of the review. An example of the system's output at the conclusion of a consultation follows:

IEP Team Participation:

Based on the information you have provided, the IEP team meeting for this student included all the participants required by the regulations.

Notification of Parent:
Based on the information you provided, the public agency's notification to the parent regarding the IEP team meeting should meet the requirement of the regulations.

Followup with Parent:
Based on the information you provided, the public agency's efforts to involve the parent in the IEP team process should meet the requirements of the regulations.

Parental Approval of Placement:
Acquisition of the parent's approval by signature for the proposed placement of the student in special education should meet the requirement of the regulations.

Development of an IEP:
The development of an IEP by the IEP team within the thirty day limitation should meet the intent of the regulations.

Components of the IEP:
The IEP has weak or missing components that are required by the regulations.

Team Approval of the IEP:
The signatures of all of the IEP team members on the IEP meets the requirement of the state rule.

Parent Approval of the IEP:

Acquisition of parental consent by signature on the IEP prior to placement and initiation of services meet the requirements of the regulations.

procedures reviewed = parent approval of IEP a (99%)
procedures reviewed = parent approval of placement a (99%)
procedures reviewed = followup with parent a (99%)
procedures reviewed = notification of parent a (99%)
procedures reviewed = components of IEP b (98%)
procedures reviewed = team approval of IEP (98%)
procedures reviewed = development of IEP a (89%)
procedures reviewed = IEP team participation a (89%)

Current Version Emphasizes Development of the IEP

Mandate Consultant is designed to review a specific portion of the IEP process, that is, the IEP team meeting and the resulting IEP for a particular case. Eventually, Mandate Consultant may be expanded to review additional portions of the process such as the referral or evaluation. However, the current version of Mandate Consultant was built with the assumption that the referral and multidisciplinary evaluation processes will be reviewed by another means. Mandate Consultant picks up the IEP process at the point in time where evaluation data has been established and an IEP team meeting is the next step.

Instructions for Operation MC

The Basics for Getting Started

Operation of MC requires two disks: (a) a special DOS diskette used to "boot" the system and (b) an M.1/EXE diskette. An additional formatted diskette is required if the user desires to make records of the consultations with MC. Following are the basic steps for getting started:

1. Insert special DOS diskette in drive A.
2. Turn microcomputer on.
3. Enter date (or just press the return key).
4. Enter time (or just press the return key).
5. Remove special DOS diskette and insert M.1/EXE diskette in drive A when this action is indicated on the screen.
6. Type "MC" when ready (do not include quotes), press the return key.
7. A banner screen will appear followed by the M.1> prompt indicating that you are in the M.1 language and are ready to run the program. Type "go" followed by a period. (From this point on all of your responses should be followed by a period.
8. Respond to the items as they are displayed on the screen. The program is ready for your input when the >> prompt appears. Sometimes there are delays of several seconds as the program accepts the user's input and generates appropriate responses or subsequent questions.
9. At the conclusion of the consultation, the summary statements will appear as the screen display

followed by the prompt "M.1>". To do subsequent consultations, the user should turn off the power and began again with step 1.

A. Typing Errors When Responding to Questions

Acceptable responses for each question appear in parenthesis after a question presented by the system, for example, (yes or no). If you notice that you made an error as you typed in a response, you can use the left arrow key to erase the error. Then type in the correct response. Remember, all responses must be followed by a period.

If you type in a response other than one of the choices in the parenthesis, follow it by a period, and press return; a message appears on the screen indicating that your response is unacceptable. Acceptable responses for this question are listed and the user is instructed to try again.

B. Following User Responses With a Period

Whenever the user types in a response to a question, the response must be followed by a period. For example, "yes.". Then the user presses the return key. If the user types in a response but fails to type a period and presses the return key; the >> prompt reappears and waits for the user to type a period and again press the return key, that is, >> "." followed by pressing the return key.

C. Auto-completion Feature for Entering User Answers

The questions presented by the system are followed by parenthesis identifying acceptable responses for the question. Because of an auto-completion feature of the system, the user only needs to type the first letter of the response and follow it by a period. For example, "y." would be the equivalent response of "yes."

The auto-completion feature has one exception: If you select the response "unknown", you must type in the entire word, that is, "unknown."

D. Key Documents from Case Files for Answering Questions Presented by MC

While the developer of Mandate Consultant anticipates that proposed users will have some previous knowledge of cases reviewed by Mandate Consultant this is not essential for successful operation of the system. In general, the questions presented by Mandate Consultant can be answered as the user reads selected items from a student's case file. Key items from a case file include: (a) the IEP, (b) the parent permission for placement form, (c) the notification to the parent of the IEP team meeting, or (d) the record of contacts with the parent and others.

E. A Record of the MC Consultation

The first question presented by the system to user states. "Do you wish to make a record of this consultation?" A "yes" response requires the user to place a formatted disk in drive B and to give a "name" to the record so it can be accessed for future use. The record name should allow the user to clearly identify the record of the consultation with its respective case file. Generally, the user will want to assign a record name that closely corresponds with the case name, (eg., johnb or jonesk) or perhaps a number that corresponds with a case numbering system (eg., case176 or 18003). Whatever system is implemented by the user, the record name or number must not have any spaces between the characters, it must not exceed eight characters in length, and it should be typed in lower case.

When the user completes the consultation, a record of that consultation referred to as a text file is created on the formatted disk in Drive B. This record or text file contains all of the questions presented by the system, the user's responses, and the conclusions by the system.

F. A Printed Copy of the MC Consultation

Using the record of the consultation that was created on the formatted disk, it is possible to print a copy of the consultation. The copy is printed using a basic word processing package with the text file on the formatted disk. For example, the user might select the PRINT command from the Wordstar (word processing program) menu. Then with the formatted disk containing the records in the disk drive, the user types in the name of the record to be printed. The specific steps for printing a copy of a record will depend on the particular word processing package available to the user. For additional directions on printing text files, refer to the documentation for your word processing program.

G. Certainty Factors

Often, rules or outcomes specific to a subject domain can not be stated with complete certainty. Thus, MC implements a weighted procedure based on the levels of uncertainty associated with specific rules or outcomes. These weighted procedures result in "certainty factors," another key element of an expert system. A certainty factor, usually based on a scale of 0-100, indicates the level of confidence that can be associated with a specific rule or outcome. For example, a certainty factor of 30 would indicate a relatively low level of confidence in the rule or outcome, whereas a certainty factor of 80 indicates a relatively high level of confidence. Examples of certainty factors appear with the section entitled Example of Conclusions Produced by MC. The certainty factors are expressed as percentages and appear in parenthesis. The first certainty factor of 99% indicates a high level of confidence in the summary statement regarding parent approval of the IEP, the second certainty factor of 99% indicates a high level of confidence in the summary statement about parent approval of placement, and so on.

Appendix U

Test Case Data Collection Form

Test Case Data Collection Form

Name of Case:____

A. Your Impressions of the Case (Note particular gaps between case file data and appropriate procedures based on federal and state regulations):

B. Differences Between Your Impressions and Conclusions from Consultation (After you run a case, note particular differences between your impressions and the system results):

C. Related Comments (Note anything about the consultation that may be relevant for future users of the system):

Appendix V

Product Prototype Questionnaire

QUESTIONNAIRE: PART A - ACCURACY

For Product Prototype Phase

After using Mandate Consultant with the set of test cases, please respond to the following questions:

1. Did you find the questions presented by the system during the consultations clear? If not, which questions were unclear?

2. Did you find the responses presented by the system during the consultation clear? If not, which responses were unclear?

3. Did you find the explanations presented by the system in response to the "why" command clear? If not, which explanations were unclear?

4. Did you find the conclusions presented by the system at the end of the consultations clear? If not, which conclusions were unclear?

5. Did the conclusions presented by the system for each case agree with your interpretation of federal and state regulations? If not, which conclusions did not agree?

QUESTIONNAIRE PART B - OPERATION

1. Did you find the instructions and documentation for the system logically organized? If not, which parts were disorganized?

2. Did you find the instructions and documentation for the system comprehensive? If not, which parts were incomplete?

3. Did you find the instructions and documentation for the system clear? If not, which parts were unclear?

4. Could you successfully operate the system based on the information in the instructions and documentation? If not, what changes would you suggest?

5. During the consultations did you find that the system's rate of response was sufficient (i.e., Was the response time fast enough so the user will not be inconvenienced, distracted, or frustrated?

6. In your opinion, would the cost of the hardware required to run this expert system make it accessible for the proposed users?

7. In your opinion, would the probable physical location (i.e., school office) of the hardware make it accessible for the proposed users?

8. In your opinion, does the time required to run the expert system make its use feasible for the proposed users?

Appendix W

Validity and Reliability of the Questionnaire for the Product Prototype Phase

Validity and Reliability of the Questionnaire for the Product Prototype Phase

Raw data from content specialists:

Item #		Rater A	Rater B
	1	4	4
	2	4	4
	3	4	4
	4	4	4
	5	4	4
	6	4	4
	7	4	4
	8	4	4
	9	4	4
	10	4	4
	11	4	4
	12	4	4
	13	4	4

Rating Scale

1 = not relevant
2 = somewhat relevant
3 = quite relevant
4 = very relevant

Coding of data using Martuza (1977) technique:

		Rater A		
		1 or 2	3 or 4	Totals
Rater B	1 or 2	0	0	0
	3 or 4	0	13	13
	Totals	0	13	13

Data analysis and results based on Martuza (1977) technique:

Interrater Agreement: Proportion of items given rating of (1 or 2) or (3 or 4) by both raters

$$\frac{\text{(Joint 1 or 2) + (Joint 3 or 4)}}{\text{Total items}} = \frac{0 + 13}{13} = 1.0$$

Content Validity Index: Proportion of items given a rating of 3 or 4 by both raters

$$\frac{\text{Joint 3 or 4}}{\text{Total items}} = \frac{13}{13} = 1.0$$

Appendix X

Sample Letter

to Cache County School District Special Educators

Dear Special Educator,

I appreciate your willingness to provide case file documentation for completion of a research grant funded to the Developmental Center for the Handicapped by the U.S. Department of Education. The case file information you are asked to provide will be used to evaluate the "knowledge" of a computer program designed to review some aspects of IEP development. I assure you that the study is not an evaluation of your skill in developing IEPs. In fact, some "weaker" IEPs would strengthen the evaluation by examining the depth of "knowledge" in the computer program.

Please randomly select one case from your classroom of students. Then make copies of the following items (or similar items that may provide the information requested):

1) the student's current IEP

2) the parent permission for placement form

3) the record of contacts with the parent or others

4) a copy of the notification to the parent of the most recent IEP team meeting

5) A form that indicates the date when the most recent evaluation of the student was completed

I recognize that some files will not include all of these items. Simply make copies of those items which are currently in the student's file.

Finally, remove all personally identifiable data about the student from the copies. I suggest you use a blue or black felt marker and draw a bold line through items such as the student's name and birthdate. I need the student's age, so write the student's age in years near the space where the birthdate usually appears.

I will plan to stop by your classroom on Monday, December 16, 1985 to pick up the case file information.

Thanks again for your assistance!

Sincerely,

James D. Parry

dh

Appendix Y

Letter/Directions for Human Experts of Phase I

Dear

I appreciated your positive response to my earlier request to participate in a research project at USU to evaluate a microcomputer-based expert system for special education. As mentioned in a previous letter, this computer program is designed to review the procedures implemented for the development of an IEP. Enclosed are materials you need to complete your part of the evaluation. Following are the directions to accompany the materials.

Directions for Reviewing Ten Case Files

The attached materials include items from ten special education [illegible] files. Please review the items from a case file and [illegible]te a report form for each case. The items for each case only provide information for a portion of the IEP process. For example, there is basically no information regarding referral evaluation. This information was intentionally withheld. Because the entire IEP process is complex, the scope of this study is limited to a portion of that process. Therefore, please focus your attention on the IEP team meeting and the resulting IEP. You are to assume that referral and evaluation have been completed appropriately and that the IEP will be implemented appropriately.

From the point in time when the IEP team meeting is scheduled to the point when the IEP is written and approved, school personnel should follow numerous procedures based on the federal and state regulations. Items such as the IEP from a student file can often provide some written evidence indicating whether appropriate procedures were implemented. For example, do the signatures on the IEP indicate that the IEP team included persons required by federal and state regulations? Or does the IEP state "annual goals including short term objectives" as required by the regulations? If a parent has not attended the IEP team meeting, is there written evidence in the case file (e.g. the Record on Contacts or other notification to the parent) indicating school personnel's attempts to involve the parent as required by regulations? Or does the notice to the parent about the IEP team meeting include information about the time, location, purpose, and who would be in attendance at the meeting as required by the regulations? As these questions would suggest, you job is to judge the procedures implemented as thoroughly as possible based on the evidence from a few case file items.

BEST COPY AVAILABLE

Sample Letter
Page 2

Directions for Completing Report Forms for Ten Cases

After the review of each case file, record your comments on a report form for the case. In the first section of the report form entitled "Summary Statements," please make several statements that summarize your perceptions of the case. Particularly note gaps between the evidence in the case file and appropriate procedures based on federal and state regualtions. For example, you might state, "Based on the evidence, the IEP meeting for this student does not include all the participants required by the regulations." You may also note consistency between the evidence and appropriate procedures. For example, you might state, "Based on the evidence, the public agency's efforts to involve the parent in the IEP team process should meet the requirements of the regulations."

In the second section of the report form entitled "Related Comments and/or Specific Citations," please make comments and/or perhaps note federal or state regulations that support your "Summary Statements" indicating gaps between appropriate procedures and evidence in the case file. For example, if you have noted in your "Summary Statements" that the IEP team did not include all the participants required, now you might make a specific statement about who was not included: "Federal Regulations 300.344 and 345 along with Utah State Rules III.B. and III.C. require that the parent participate in the IEP team meetings." Or, you might state, "Without the parent, the IEP team is inconsistent with the intent of the regulations."

Directions for Informed Consent Form

Included with this packet is an Informed Consent Form requiring your signature. The form describes your involvement in this research activity. Please sign the form and return with the other materials.

General Information Form

Please take a few moments and complete the General Information Form included with this packet.

Your prompt response to this packet would be greatly appreciated. If you have questions, please call me (Office

753-7973 or Home 753-5742). Upon completion, return the ten report forms, the Informed Consent Form, and the General Information Form to:

Jim Parry
Systems Impact
UMC-6810
Utah State University
Logan, UT 84322

You need not return the ten case files.

Thanks again for your help!

Sincerely,

James D. Parry

Appendix Z

Cumulative File Report Form

Cumulative File Report Form

(See directions provided in letter)

Name of Case: ______________________________

A. Summary Statements:

B. Related Comments and/or Specific Citations:

Appendix AA

Informed Consent Form for Human Experts

SUGGESTION:

As you begin this review process, it may be helpful to think of yourself as the person responsible for approval of these particular case files. What information in the case file suggests that appropriate procedures were not followed for the IEP team meeting and the development of the IEP? Or, before you would approve this case as in compliance with federal and state regulations, what weaknesses or problems need to be addressed?

Informed Consent Form

Study: The Development and Validation of a Computer-Based Expert System for Examining the Implementation of Special Education Regulations.

Study Administration: Dr. Alan Hofmeister
James D. Parry

Any questions regarding this form should be directed to James D. Parry, UMC-6800, Utah State University, Logan, UT 84322.

As a subject of the above referenced study, I agree to review the information in selected special education case files and complete a brief summary report for each case. I understand that my reports will be used in the validation component of the study.

While the contents of the summary reports may be stated in the results of the study, neither my name nor other personally identifiable data will appear in the study results. In addition, a copy of the results of the study will be provided to me upon the study's conclusion.

Signature of Subject

Please return signed form to James D. Parry

Appendix BB

General Information Form for Human Experts

General Information Form

Name:______________________________

Title:_____________________________

Social Security Number:__________________

1. Total years of experience in special education (include classroom and administrative experience):________

2. Have you participated in any due process hearing officer training?

   ___Yes ___No

   If yes, have you been appointed as a hearing officer?

   ___Yes ___No

3. Have you participated in any special training regarding appropriate procedures for implementation of the IEP process?

   ___Yes ___No

   If yes, name at least one of these training sessions:

Appendix CC

Suggestion Notice

Appendix DD

Letter/Directions for Evaluators of Phase II

Dear

I appreciated your positive response to my earlier request to participate in this research project. As you recall, I specifically need your assistance in an evaluation activity that is part of a "double-blinded" study. The following paragraphs provide the directions for your involvement.

Materials contained in this packet:

A) Informed Consent Form
B) General Information Form
C) Ten pocket folders that each contain
    1) A brief case summary
    2) Seven case file reports with attached rating forms
D) Copy of the directions provided to the reviewers who wrote the case file reports for the ten case files.
E) Copy of "Suggestion" Notice

A colored pocket folder has been prepared for each of the ten special education cases. On the left side of the folder is a brief case file for you to review. On the right side of the folder you will find seven case file reports that have been completed by seven different reviewers. A rating form for the "blinded" evaluation has been attached to the front of each of the seven case file reports. Your assignment is to: (a) review each brief case file (found on the left side of the pocket folder) and formulate your own judgement regarding compliance concerns, and (b) complete the Rating Forms attached to the seven case file reports (found on the right side of the pocket folder).

As you prepare for this task, you are encouraged to read the directions that were provided to the reviewers that previously reviewed the brief case files and wrote the case file reports. A copy of these directions is enclosed. In addition, a "Suggestion" notice was attached to the brief case files that were provided to the reviewers. The notice briefly restated the reviewers' assignment in an effort to keep them on task. A copy of the "Suggestion" notice is also enclosed for your information.

Your prompt response to this packet would be greatly appreciated. As mentioned in my earlier correspondence, I hope you are able to complete the task and return the results to me by February 15, 1986. If you have questions, please call me (Office 753-7973 or Home 753-5742). Upon completion, return the seven Rating Forms for each case file, the Informed Consent Form, and the General Information Form to:

Jim Parry
Systems Impact
UMC-6810
Utah State University
Logan, UT 84322

Thanks again for your assistance!

Sincerely,

James D. Parry

Enclosures

Appendix EE

Informed Consent Form for Evaluators

Informed Consent Form

Study: A Study of the Implementation of Special Education Regulations

Study Administration: Dr. Alan Hofmeister
and James D. Parry

Any questions regarding this form should be directed to James D. Parry, UMC-6800, Utah State University, Logan, UT 84322.

As a subject of the above referenced study, I agree to review the brief case files of ten special education cases and then rate seven brief summary reports for each case. I understand that my ratings will be used in the evaluation component of a "double-blinded" study.

While the contents of the rating forms may be stated in the results of the study, neither my name nor other personally identifiable data will appear in the study results. In addition, a copy of the results of the study will be provided to me upon the study's conclusion.

____
Signature of Subject

Please return signed form to James D. Parry

Appendix FF

General Information Form for Evaluators

General Information Form

Name:______________________________

Title:______________________________

Social Security Number:__________________

1. Total years of experience in special education (include classroom and administrative experience):________

2. Have you participated in any due process hearing officer training?

   ___Yes ___No

   If yes, were you a trainer or trainee?

   ___Yes ___No

3. Have you been appointed as a hearing officer?

   ___Yes ___No

4. Have you participated in any special training regarding appropriate procedures for implementation of the IEP process?

   ___Yes ___No

   If yes, name at least one of these training sessions. Also, describe your role; that is, trainer, trainee, or both.

Appendix GG

Rating Form for Phase II

Rating Form for Phase II Evaluation

Please indicate your opinion regarding how well the content of the attached report identifies gaps between the procedures contained in the federal and state regulations and the procedures followed for developing the IEP of a special education student as documented in the brief case file. In brief, how well does the report address compliance concerns:

☐ Ideal: The information summarized in the report is synonymous with what I would have written.

☐ Acceptable: The information summarized in the report differs from what I would have written, but it is acceptable.

☐ Less than acceptable: The information summarized in the the report is inaccurate and/or inadequate, however, I would consider these deficiencies minor. Please identify specific deficiencies:

☐ Unacceptable: The information summarized in the report is inaccurate and/or inadequate, and I would consider these deficiencies major. Please identify specific deficiencies:

Appendix HH

Expert Conclusions Coded for Each Case by Special Education Graduate Student

228

Expert Conclusions Coded for Each Case by Special Education Graduate Student

B = conclusion by MC and HE
M = conclusion only noted by MC
H = conclusion only noted by HE
D = disagreement between experts

ALAN	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
IEP not developed in 30 days	B		B				
Measurable, short term objectives not included	B	B	B			B	B
Does not specify all related services	M						
No schedule for evaluating short term objectives	M						
No parental consent prior to placement and initiation of services	B		B	B	B		B
Total Number of Interexpert Agreements	9	4	9	4	4	4	8

BRAD	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
Does not state all related services and appropriate goals	B		B			B	
Does not specify all service providers	B				B		
No schedule for evaluating short term objectives	B			B			
Time in regular and sp. ed. not changed			H	H			H
No permission to place for previous placement			H				
No procedures for evaluating objectives				H			
No date for initiating services				H			
No date for 3-year evaluation					H		
Total Number of Interexpert Agreements	4	0	4	3	1	2	2

CRAIG	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	B		B				
No parental consent prior to initiation of services	B		B	B	B		B
All service providers must be specified	B					B	
No date for initiation of services	B			B		B	
No objective criteria	B	B				B	
No evaluation procedures	B	B		B			
No schedule for evaluating short term objectives	B			B	B		
No parental approval of proposed placement	B		B	B	B		
Time unit not specified (per day, week, etc.,)		H					
No evidence of considering LRE			H				H
No projected date for 3-year re-evaluation					H		
P.E. must be addressed as to amount of time and type of service							H
Total Number of Interexpert Agreements	17	5	9	12	9	5	5

JOE	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
IEP team did not include all required participants	B		B	B	B	B	B
Student could have attended IEP meeting	B					B	
Short term, measurable objectives not included	B			B			
Does not indicate duration of services	B		B	B		B	
No schedule for evaluating short term objectives	M						
Participants must sign, date, and indicate position	B		B		B	B	B
Does not indicate time unit (per day, week, etc.)		H					
No justification for not placing in a LRE		H		H			H
Some goals not related to weaknesses		H					H
Total Number of Interexpert Agreements	14	3	12	11	9	13	12

MARGARET	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
IEP not written within 30 days	M/D		H/D				
Parental consent not obtained prior to placement	B		B	B	B		
Measurable, short term objectives not included	B		B	B		B	
Does not specify all related services	M/D		H/D				
No schedule for evaluating short term objectives	M						
Time in sp. ed. not specified		H					
Date of initiation of services not specified				H			
Total Number of Interexpert Agreements	6	0	6	6	3	3	0

MARYANN	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
Measurable, short term objectives not included	M/D		H/D				
No projected dates for init. of services	B			B			
No projected dates for duration of services	B		B			B	B
No objective criteria	M						
Evaluation procedures not specified	M						
No schedule for evaluating short term objectives	M						
Time unit not specified (per day, week, etc.)		H					
Initial placement made prior to parent authorization or IEP			H	H			
Total Number of Interexpert Agreements	4	0	2	2	0	3	3

RICHARD	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
Service providers not identified	B		B				
No projected dates for initiation of services	B		B	B	B	B	
Does not state objective crit.	B		B	B		B	
Does not state eval. procedures	B	B		B	B		
IEP not signed prior to placement	B		B				
Total time in special ed. not specified			H				
IEP review not within one year			H	H			
Goals set in area with no data to indicate need for them							H
Time unit not specified (per day, week, etc.)		H					
Total Number of Interexpert Agreements	12	3	9	10	7	7	0

ROBERT	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
No projected dates for duration of services	M						
No schedule for eval. short term objectives	M						
IEP not signed prior to placement	B						B
Anticipated length of service not specified			H	H		H	
IEP review not within one year			H	H			
IEP not completed in 30 days			H				
No date for initiation of services				H			
No goals and obj. for P.E.							H
Time in regular ed. not specified at re-evaluation						H	H
Total Number of Interexpert Agreements	1	0	3	3	0	1	2

RUSSELL	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
Time in regular ed. not specified	B/D		H/D	B/D			B/D
No projected dates for initiation of services	B			B			
No objective criteria	B		B				B
No schedule for evaluating short term objectives	B			B			
No evaluation procedures	B			B			
No goals related to some stated weaknesses		H					
Time unit not specified (per day, week, etc.)		H					
Level of performance not based on formal/informal measures				H			H
Total Number of Interexpert Agreements	7	0	1	5	0	0	4

SHERRI	MC	HE1	HE2	HE3	HE4	HE5	HE6
Parents understood proceedings	M						
No schedule for evaluating short term objectives	B		B	B			
Inadequate present level of performance				H			H
No measurable short term obj.				H			
Time unit not specified (per day, week, etc.)		H		H			
Total Number of Interexpert Agreements	2	1	2	3	0	0	1

Appendix II

Independent Evaluator Ratings for Each Expert's Conclusions

Evaluator Ratings of Mandate Consultant Conclusions by Case

	Eval 1				Eval 2				Eval 3			
	I	A	L	U	I	A	L	U	I	A	L	U
Alan			X				X				X	
Brad		X				X					X	
Craig		X				X				X		
Joe		X					X				X	
Margaret		X				X				X		
Maryann		X					X					X
Richard		X			X					X		
Robert		X				X						X
Russell			X			X				X		
Sherri			X				X					X
	0	7	3	0	1	5	4	0	0	4	3	3

Eval = Evaluator

I = Ideal
A = Acceptable
L = Less than Acceptable
U = Unacceptable

Evaluator Ratings of Human Expert 1 Conclusions by Case

	Eval 1				Eval 2				Eval 3			
	I	A	L	U	I	A	L	U	I	A	L	U
Alan			X				X					X
Brad			X			X						X
Craig			X					X				X
Joe				X				X				X
Margaret			X				X					X
Maryann			X				X					X
Richard			X			X					X	
Robert			X			X						X
Russell			X			X						X
Sherri			X				X					X
	0	0	9	1	0	4	4	2	0	0	1	9

Eval = Evaluator

I = Ideal
A = Acceptable
L = Less than Acceptable
U = Unacceptable

Evaluator Ratings of Human Expert 2 Conclusions by Case

	Eval 1				Eval 2				Eval 3			
	I	A	L	U	I	A	L	U	I	A	L	U
Alan		X					X			X		
Brad		X				X				X		
Craig	X						X				X	
Joe		X				X					X	
Margaret		X					X			X		
Maryann			X			X					X	
Richard		X				X					X	
Robert		X				X				X		
Russell			X			X					X	
Sherri		X					X				X	
	1	7	2	0	0	6	4	0	0	4	6	0

Eval = Evaluator

I = Ideal
A = Acceptable
L = Less than Acceptable
U = Unacceptable

Evaluator Ratings of Human Expert 3 Conclusions by Case

	Eval 1				Eval 2				Eval 3			
	I	A	L	U	I	A	L	U	I	A	L	U
Alan			X				X				X	
Brad			X			X				X		
Craig			X				X		X			
Joe	X						X				X	
Margaret		X					X					X
Maryann			X			X					X	
Richard		X					X			X		
Robert		X				X				X		
Russell			X				X			X		
Sherri			X				X			X		
	1	3	6	0	0	3	7	0	1	5	3	1

Eval = Evaluator

I = Ideal
A = Acceptable
L = Less than Acceptable
U = Unacceptable

Evaluator Ratings of Human Expert 4 Conclusions by Case

	Eval 1				Eval 2				Eval 3			
	I	A	L	U	I	A	L	U	I	A	L	U
Alan			X					X				X
Brad			X				X				X	
Craig			X				X				X	
Joe			X					X				X
Margaret			X				X					X
Maryann		X					X					X
Richard		X						X				X
Robert			X				X					X
Russell		X						X				X
Sherri			X				X					X
	0	3	7	0	0	0	6	4	0	0	2	8

Eval = Evaluator

I = Ideal
A = Acceptable
L = Less than Acceptable
U = Unacceptable

Evaluator Ratings of Human Expert 5 Conclusions by Case

	Eval 1				Eval 2				Eval 3			
	I	A	L	U	I	A	L	U	I	A	L	U
Alan			X					X				X
Brad			X				X					X
Craig				X				X				X
Joe		X					X					X
Margaret			X				X					X
Maryann			X				X				X	
Richard			X				X					X
Robert				X				X				X
Russell			X				X					X
Sherri			X				X					X
	0	1	7	2	0	0	7	3	0	0	1	9

Eval = Evaluator

I = Ideal
A = Acceptable
L = Less than Acceptable
U = Unacceptable

Evaluator Ratings of Human Expert 6 Conclusions by Case

	Eval 1				Eval 2				Eval 3			
	I	A	L	U	I	A	L	U	I	A	L	U
Alan			X			X						X
Brad		X				X						X
Craig		X						X				X
Joe			X				X				X	
Margaret			X				X					X
Maryann			X			X						X
Richard			X				X					X
Robert		X				X						X
Russell			X			X						X
Sherri			X				X					X
	0	3	7	0	0	5	4	1	0	0	1	9

Eval = Evaluator

I = Ideal
A = Acceptable
L = Less than Acceptable
U = Unacceptable